

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 6, 2003, 10:00:48 ; Search time 256 Seconds
(without alignments)
3063.765 Million cell updates/sec

Title: US-09-856-836-2
Sequence: 1 MAARKQIPUCSGHTRVYDX.....EELASNSDSTSTPEVKA 351

Scoring table:
BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ -p2n.model -DEV-xlh
-O/-cgn2_1/USPFO.spool/OS0856836/runal.31012003.144236.22173/app.query.fasta_1.519
-DB-N-Geneseq.101002 -OEMT-fastap -SUPFIX-rng -MINMATCH=0.1 -LOOPC=0
-LIST=5 -DOCALL=200 -THR-SCORE=500 -MINLEN=0 -MAXLEN=200000000
-MODE-LOCAL -OUTFMT=plc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER-US09856836.ecgn_1.1.0.etrnal.31012003.144236.22173 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N_Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1832	99.6	1630	21	AAA49249
2	1779.5	96.7	1820	20	AAV59096
3	1758.5	95.6	1847	19	AAV54162
4	1737	94.4	2812	23	ABV22737
5	1737	94.4	2812	23	ABV28565
6	1347.5	73.2	1430	21	AACT8126
7	1334.5	72.5	15320	22	AAK78660
8	1328.5	72.2	15332	22	AAK78662
9	1052.5	57.2	1331	23	ABL06979
10	876.5	47.6	4891	23	ABL06978
11	803.5	43.7	4179	23	ABL09294
12	789.5	42.9	1753	21	AAAC43745
13	781.5	42.5	1410	21	AAAC47397
14	771	41.9	1540	21	AAAC47955
15	770	41.8	1538	21	AAAC46123
16	759.5	41.3	1410	21	AAAC38458
17	615	33.4	522	24	ABO58251
18	517.5	28.1	313	23	AAAS58064
19	467	26.5	311	24	AAKS3945
20	454.5	24.7	2901	23	ABL16514
21	445	24.2	469	24	ABL93455
22	427.5	23.2	901	23	ABL16515
23	426.5	23.2	3165	23	ABL03412
24	425	23.1	686	20	AAZ24536
25	425	23.1	686	21	AAAC65775
26	425	23.1	686	24	ABL48994
27	424.5	23.1	1072	21	AAAC3747
28	368.5	20.0	1246	21	AAAC3847
29	368.5	20.0	1253	21	AAAC3946
30	368.5	20.0	1256	21	AAAC3404
31	357.5	19.4	1276	21	AAAC47884
32	356.5	19.4	507	21	AAAC45273
33	353.5	19.2	987	21	AAAC43473
34	321	17.6	1166	23	ABL14459
35	321	17.4	223	15	AAO76451
36	301	16.4	3398	23	ABL14458
37	289	16.2	1710	22	AAK94176
38	279	15.2	1512	22	AAK93563
39	274	14.9	3450	22	AAA91119
40	274	14.9	3714	22	AAA91117
41	274	14.9	3747	22	AAA91115
42	274	14.9	3747	22	AAA91118
43	274	14.9	3747	22	AAA91120
44	274	14.9	3747	22	AAA91121
45	274	14.9	3747	22	AAA91122

ALIGNMENTS

RESULT 1
AAA49249 standard: cDNA: 1630 BP.

XX AAA49249;

XX 25-SEP-2000 (first entry)

DE Serine-threonine kinase receptor associated protein (STRAP) DNA sequence.

KW Serine-threonine kinase receptor associated protein; STRAP; mouse;

KW transforming growth factor beta; TGF-beta; proliferation; tumour growth;

KW cytosolic; anti-arteriosclerotic; anti-diabetic; nephrotic; cancer;

XX atherosclerosis; diabetes; ss.

XX Mus sp.

OS

XX

XX

XX

Location/Qualifiers

[illegible]

Db	439	GCCTTTGGGGTGCACACACGTAATTAAGATAGCCACCAAGCAAGCAGCTACACAGCTGCAGAT	498
OY	81	PheThrAlaIysValITrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisIlys	100
Db	499	TTTCACAGCCAAAGTGTGGGATGCGTCTCAGAGAGATGCAATTTGATGACCTCGCTCATAAA	558
OY	101	HisIleValIysThrValAspPheThrGlnAspSerAsnTyrIleu**ThrGlyGln	120
Db	559	CACATGTGCAAGACTGGTGGATTTCCACCAGAGATAGTAATTAATTTGTTAAACCGGGGACAG	618
OY	121	AspIysLeuLeuArgIleTyrAspLeuAsnIysProGluAlaGluProIysGluIleSer	140
Db	619	GATAAACGTTACCATATATGTACTTGACATTACAAACCTGAAACCGACAACTTAAGGAATTTAGT	678
OY	141	GlyHisThrSerGlyIleIysIysAlaLeuTyrPcysSerAspAspIysGlnIleLeuSer	160
Db	679	GGTCATCTCTTGCTGATATAAAAAAGCTCTGGTGGCAGTGAAGATTAACAGATTTCTTTCT	738
OY	161	AlaAspAspIysThrValArgLeuTPRAspHisAlaThrMetThrGluValIysSerLeu	180
Db	739	GCTGTGTCAAAACCTGTTCCACTTTGGACATCATCATATGACAGAAAGTGAACCTCTTA	798
OY	181	AsnProAsnMetSerValSerSerMetGluTyrIleProGluGlyIleIleuValIle	200
Db	799	AAITTTATATATGTGTGTAGTAGATAGGAATATATCTCGAGGGAGAGATTTTGGTATTA	858
OY	201	ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleIysSerPhe	220
Db	859	ACTTATGAGACATCTATTTGCTTTTCATAGTGCAGTAAAGTTTGGACCCATTTAAATCTTTT	918
OY	221	GluAlaProAlaThrIleAsnSerAlaSer**HisProGluIuIysGluPheLeuValAla	240
Db	919	GAACCTCTCGCAACCATCATTTCTGCATCTCTTCATCTCGAAGAAAGAAATTTCTGTGCA	978
OY	241	GlyGlyGluAspPheIysLeuTyrIysTyrAspTyrAsnSerGlyGluIuLeuIuSer	260
Db	979	GGCGGTGAGATTTTAACTTATTAAGTATGATTAATTAATGAGGAAAGAAATTAAGAAATCC	1038
OY	261	TyrIysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyIuLeuTyr	280
Db	1039	TACAAGGAGACCTTTGGTCTATTTCACTGTGTGAGATTAACTTCGTGATGAGCAACTCAT	1098
OY	281	AlaSerGlySerGluAspGlyThrIleuArgLeuTyrPglThrValValGlyIysThrTyr	300
Db	1099	GCCAGTGGTTCAGAAAGATGGAAACATTGACACTGTGCAAACTGTGTAGGAAAAACGTAT	1158
OY	301	GlyLeuThrIysCysVal***ProGluGluAspSerGlyGluLeuAlaIysProIysIle	320
Db	1159	GGCCTTTGGAAATTTGTGCTCTCGAAGAAAGATAGTGTGACTGTGCAAAAGCCAAAGATT	1218
OY	321	GlyPheProGluThrAlaGluGluIuLeuAlaGluGlnIleAlaSerGluAsnSerAsp	340
Db	1219	GGTTTTCACAGACAAACAGAAAGAGAGCTA---GAAGAAATTTGCTTCACAGATTCAGAT	1275
OY	341	SerIleTyrSerSerThrProGluValIysAla 351	
Db	1276	TGCATCTTTCCTTCACAGCTCCTGATGTTAAGGCC 1308	
RESULT 3			
ID	AAV54162		
TD	AAV54162 standard; cDNA; 1847 BP.		
XX	AAV54162;		
XX	AC		
XX	21-DEC-1998 (first entry)		
XX			
DE	Nucleotide sequence 7 from human cell line SH-SY5Y.		
XX			
KW	Human; SH-SY5Y; apoptosis; antibody; immunohistological staining;		
KW	inhibition; ds.		
XX			
OS	Homo sapiens.		
XX			

FH Key Location/Qualifiers
 FT CDS 408..1103
 FT /tag= a
 FT /product= "SH-SY5Y protein"
 XX WO9839437-A1.
 XX PN 11-SEP-1998.
 XX PD 05-MAR-1998; 98WO-JP00905.
 XX PF 05-MAR-1998; 97JP-0050302.
 XX PR 05-MAR-1997; 97JP-0050302.
 XX PA (KYOM) KYOMA HAKKO KOGYO KK.
 XX PI Sakaki Y;
 XX DR WPI: 1998-495844/42.
 XX DR P-PSDB: AAM54161.
 XX PT Novel apoptosis-related DNAs and proteins - for diagnosis,
 XX preventing or treating diseases associated with apoptosis.
 XX PS Claim 1: Pages 40-43; 70pp; Japanese.
 CC This is the nucleotide sequence of a nucleic acid of the human cell
 CC line SH-SY5Y, used in the method of the invention involving the use of
 CC novel apoptosis-related DNAs and proteins. The inventions can be used
 CC as diagnostic reagents for apoptosis e.g. (monoclonal) antibodies for
 CC the protein, as a reagent in immunohistological staining, as apoptosis
 CC inhibitors. It can also be used for treatment of apoptosis-related
 CC diseases.
 XX SQ Sequence 1847 BP; 513 A; 389 C; 422 G; 523 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,92e-188 Length: 1847
 Score: 1758.50 Matches: 338
 Percent Similarity: 96.88% Conservative: 4
 Best Local Similarity: 95.75% Mismatches: 8
 Query Match: 95.57% Indels: 3
 DB: 19 Gaps: 1
 US-09-856-836-2 (1-351) x AAV54162 (1-1847)
 OY 1 MetaLametaArglnThrProLeuThrCysSerGlyHisThrArgProValAsp*** 20
 DB 295 ATGGCAATGAGACAGACGCCGCTGCTGCTGCGCACAGCGGACCGGTTGATTG 354
 OY 21 AlaPhSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 355 GCGTTCAGTGGCAGCAGCGCTTATGGGTATTTCTTAATCAGCCCTTGCAAAAGTGTAAA 414
 OY 41 PrometLeuArgln-GlyAspThrGlyAspTrpIleGlyThrPheLeuLysLysG 60
 DB 415 CCGATCTCAGCGCAGGAGATACAGAGACTGATGGTAACATTTTGGCTATAAAGG 474
 OY 60 yAlaValITrPGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAs 80
 DB 475 TGGTGTGGGGTGCACACTGATAAGATGCGACCAAGGACGTCACAGCTGCA 534
 OY 80 PheThrAlaLysValITrPAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 DB 535 TTTCACAGCCAAAGTGGGATGCTGCTCAGAGATGATGATGATGATGATGATGAT 594
 OY 100 SHISLIEValLysThrValAspPheThrGlnAspSerAspTyrLeu***ThrGlyGlyG 120
 DB 595 ACACATTTGTCACAACTGATGATTTACGCGAGTAGATTAATTAATTTGTTAACCGGGACA 654
 OY 120 nAspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluLys 140
 DB 655 GGATTAACGTGTACGATATATGACTTGAACAACCTGGAAGGCAACCTTAAGGAATTAAG 714

OY 140 rGlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSe 160
 DB 715 TGGTCACTACTTTCGTATATAAAAAAGCTCTGCTGCGACGAGGATTAACAGATTCTTC 774
 OY 160 rAlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLe 180
 DB 775 TGGTGTAGCAAAACTGTTGACTTGGGATCATGCTACTATGACAGAACTGAATCTCT 834
 OY 180 uAsnPheAsnMetSerValSerSerMetGluTyrIleProGluGluValIleLeuVal 200
 DB 835 AAATTTAATATGCTGTTAGTAGATGAGATATATCTTGAGGAGACATTTGGTTAT 894
 OY 200 eThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSerPh 220
 DB 895 AACTTATGACAGATCATCTCTTTCATAGTCGAGTAGATTGGACCAATTAATCTCT 954
 OY 220 eGluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuVal 240
 DB 955 TGAAGCTCCGCAACCATCATCTGCTATCTCTCATCTGAGAAAGAAATTTCTGTGC 1014
 OY 240 aGlyGlyGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSe 260
 DB 1015 AGCGCGTGAAGATTTTAACTTTATAGTATGATTTAATATAGTGAGACAAATTAAGATC 1074
 OY 260 rTyrLysGlyHisPheGlyProIleHisCysValArgPheSer- ProAspGlyGluLeu 280
 DB 1075 CTCAAGAGGACACTTGGCTCTCTATTCACGTGTGAGATTTAGTCCCTGATGAGAACTCT 1134
 OY 280 yAlaAspSerGlySerGluAspGlyThrLeuArgLeuTrpGlnThrValGlyLysThr 300
 DB 1135 ATCCCACTGGTTCAGAAAGATGGAACATTTGAGCTATGCGCAACTGTGTGAAAAACGT 1194
 OY 300 yGlyLeuTrpLysCysVal***ProGluGluAspSerGlyGluLeuAlaLysProLys 320
 DB 1195 ATGGCCCTTGGAAATGCTGCTCTCTCAAGACATATGCTGAGCTGCAACCAAGAGA 1254
 OY 320 leGlyPheProGluThrAlaGluGluGluLeuAlaGluIleAlaSerGluAsnSer 340
 DB 1255 TTGGTTTTCAGAGACAAACAGACAGAGCTA--GAAGAAATGCTTCAGAGAAATTCAG 1311
 OY 340 sPseIleTyrSerSerThrProGluValLysAla 351
 DB 1312 ATTGCATCTTCTCTCAGCTGCTGATGTTAAGGC 1346
 RESULT 4
 ABV22737
 ID ABV22737 standard; cDNA; 2812 BP.
 XX AC ABV22737;
 XX DT 13-SEP-2002 (first entry)
 XX DE Human prostate expression marker cDNA 22728.
 XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 XX KW pharmacogenomic marker; gene; ss.
 XX OS Homo sapiens.
 XX PN W0200160860-A2.
 XX PD 23-AUG-2001.
 XX PF 20-FEB-2001; 2001WO-US05171.
 XX PR 17-FEB-2000; 2000US-183319P.
 XX PR 16-MAR-2000; 2000US-189862P.
 XX PR 25-MAY-2000; 2000US-207454P.
 XX PR 09-JUN-2000; 2000US-211314P.
 XX PR 18-JUL-2000; 2000US-219007P.
 XX PR 13-DEC-2000; 2000US-255281P.
 XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 5974-5975; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV6213) of the specification or its complement. (I) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate cell carcinogenic potential of a compound;

(g) determining whether prostate cancer has metastasized in a patient;

(h) assessing the aggressiveness or indolence of prostate cancer in a patient;

(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 2812 BP; 778 A; 614 C; 601 G; 819 T; 0 other;

Alignment Scores:

Pred. No.:	1.43e-185	Length:	2812
Score:	1737.00	Matches:	336
Percent Similarity:	96.33%	Conservative:	5
Best Local Similarity:	94.92%	Mismatches:	8
Query Match:	94.40%	Indels:	5
DB:	23	Gaps:	1

US-09-856-836-2 (1-351) x ABV28565 (1-2812)

QY 1 MetAlaMeArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
 DB 336 ATGGCAATGACAGACAGCGCTCAGCTGCGCACCGACCGACCGCTGGTGTGTTGG 395
 QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 396 GCCTTCAGTGCATCAGCGCTATGCTTATCTTATCAGCGCTGCAAAAGTGTAA 455
 QY 41 ProMetLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 DB 456 CCTAATGCTACGCGCAGGAGATACAGAGAGCTGATGGAACATTTTGGCTCATAAAGGT 515
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrLysAlaAlaAsp 80
 DB 516 GGTGTTGGGGTGCACACCTGAATAGGATGCCACCAAGCAGCTACAGCAGCTGCAGAT 575
 QY 81 PheThrAlaLysValIleTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 DB 576 TTCACAGCCAAAGTGGGATGCTGCTCAGAGATGATGATGACCTGGCTCATAA 635
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAspTyrLeu***ThrGlyGlyGln 120
 DB 636 CACATTTGCAACACCTGGATTCACCGACAGATGAATTAATTTGTTAAACGGGGACAG 695
 QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluLysSer 140
 DB 696 GATTAACCTGTACGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 755
 QY 141 GlyHisThrSer-GlyTyrLysLysAlaLeuTyrCysSer-AspAspLysGln-IleLeu 159
 DB 756 GGTCACTACTCTGGGTATTAATAAAAGCTGCTGTGTGTCAGTGTGATTAACAAATTCCT 815
 QY 160 SerAlaAspAspLysThrValArgLeuThrPheHisAlaThrMetThrGlyValLysSer 179
 DB 816 TCTGCTGATGACAAACCTGTCACCTTGGGATGATGATGATGATGATGATGATGATGAT 875
 QY 180 LeuAsnPheAsnMetSerValSerSerMetGluTyrIleProGluLysGluIleLeuVal 199
 DB 876 CTAAATTTTAATATATGCTGTAGTACTATGATGATGATGATGATGATGATGATGATGAT 955

QY 200 IleThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSer 219
 DB 936 ATAACTTATGAGCATCTATCTCTTCTTCTATGATGATGATGATGATGATGATGATGATGAT 995
 QY 220 PheGluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuVal 239
 DB 996 TTGGAAGCTCCCGCAACCATCAATCTGCTCTTCATCCCGAGAGAAAGAAATTTCTGT 1055
 QY 240 AlaGlyGlyGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyLysGluLeuGlu 259
 DB 1056 GAGGCGGGGAAAGATTTTAATCTTATATGATGATGATGATGATGATGATGATGATGATGAT 1115
 QY 260 SerTyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeu 279
 DB 1116 TCCTACAAAGGACACTTGTGCTCTTACTGATGATGATGATGATGATGATGATGATGATGAT 1175
 QY 280 TyrAlaSerGlySerGluAspGlyThrLeuAlaGlyLeuTrpGlnThrValAlaGlyLysThr 299
 DB 1176 TATGCCAGTGTTCAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 1235
 QY 300 TyrGlyLeuTrpLysCysVal***ProGluLysAspSerGlyLysLeuAlaLysProLys 319
 DB 1236 TATGCTTGTGAAATGTGTGCTCTCTGAGAAAGATGATGATGATGATGATGATGATGATGAT 1295
 QY 320 IleGlyPheProGluThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluAsnSer 339
 DB 1296 ATTTGTTTCCAGACAGACAG 1349
 QY 340 AspSerIleTyrSerSerThrProGluValLysAla 351
 DB 1350 GATTCATCT 1385
 RESULT 6
 AAC78126
 ID AAC78126 standard; cDNA; 1430 BP.
 AC AAC78126;
 XX 08-FEB-2001 (first entry)
 DE Human cancer associated gene sequence SEQ ID NO:520.
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antiaslmetabolic; antirheumatic; antiallergic; antiviral;
 KW antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
 KW vasotropic; antipsoriatic; antianogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 OS Homo sapiens.
 PN WO20005350-A1.
 XX 21-SEP-2000.
 PD 08-MAR-2000; 2000WO-US05882.
 PF 12-MAR-1999; 99US-0124270.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI MPI, 2000-587533/55.
 DR P-PSDB: AAB3917.
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -

PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249309.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0239678.
 PA (HDMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483426/52.
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating cancers and
 XX metastasis -
 PS Disclosure; SEQ ID NO 33472; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 15320 BP; 4488 A; 2687 C; 3177 G; 4968 T; 0 other;

Alignment Scores:

Pred. No.:	4,43e-139	Length:	15320
Score:	1334.50	Matches:	291
Percent Similarity:	66.81%	Conservative:	13
Best Local Similarity:	63.96%	Mismatches:	44
Query Match:	72.53%	Indels:	110
DB:	22	Gaps:	3

US-09-856-836-2 (1-351) x AAK78660 (1-15320)

QY	1	MetAlaMeTArgSlnIhnrProLeuThrCysSerGlyHisThrArgProValValAsp***	20
Db	13430	ATPAGCAATAGAGACTGATGCTGCTCACCTCTTCTGGCCACATATGCGCCGTGGCTGACTTGG	13371
QY	21	AlaPheSerGlyLeuThrPro-----	27
Db	13370	GCCTTCAGTGGCATCATGCGCTTAAGGGGGCTTTTTTTCCTCAAGACAGAACTTGCTCG	13311
QY	27	-----	27
Db	13310	TGCGCCAGGCTGGAGTGCAGTGTGTCAAATCTTGCCACTGCAAACTCCGCTCCAGGT	13251
QY	27	-----	27
Db	13250	TCAAGTGATTCTTCCTGCCACAGCTCCCGAGTAGCTGGATTACAGGCGACGCTATAC	13191
QY	27	-----	27
Db	13190	TGCTGGCTAATTTTGTATTATTTTATGTAAGAGAGGGGTTCACCAATGTTGGCCAGCTGT	13131
QY	27	-----	27
Db	13130	CTCGAACTTGTGACCTCAGGTGATCCGCCCTGGCTGGCTCCCAAAAGTCTGGATTACA	13071
QY	28	-----TyrGlyYrPheLeuIleSerAlaCysIAsAs	38
Db	13070	GGCGTGAGCCACTGGCGCTGGCCAGCTCTTAATAGATATTTCTAAATCAGTGTGCAAAA	13011
QY	38	pGlyIysPromeLeuArgInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHis	58
Db	13010	CGGCAAACTATGCTATGTGCGAGGAGATACAGAGACTGGATTGGACAAATTTGGGTCA	12951
QY	58	slYsgIyAlaValITripGlyAlaThrLeuAsnIysAspAlaThrLysAlaAlaIaThrAlaI	78
Db	12950	TAAAGTGTGCTGTGGAAATGCAACACTAAATTAAGATGGCCACCAAGCAGCACAGACAG	12891
QY	78	aAlaAspPheThrAlaIysValITripAspAlaValSerGlyAspGluLeuMetThrLeuAl	98
Db	12890	TGCAAGTTTTCCAA--AACTG--TGGGATGCTATCCAGAGATGAACGTATGATACCTGGC	12835
QY	98	aHisIysHisIleValIlystThrValAspPheThrGlnAspSerAsnYrLeu**ThrGln	118
Db	12834	TCATTAACGACATATTCAGAGACTATGATTTTCATGACAGATACTAATTTATTTTGAACCG	12775
QY	118	yGlnIynAspIysLeuLeuArgIleTyrAspLeuAsnIysProGluAlaGluProIysGlnI	138
Db	12774	GGGACAGATAAAGCTGTTTGCATATATGACTTGAACAAACCCGAAGCAGACCTGAGGA	12715
QY	138	uIleSerGlyHisThrSerGlyIleIysIysAlaLeuTrpCysSerAspAspIysGlnI	158
Db	12714	AATTTATGATGATCCACACTCGATATTTAAAAAGCCCTATGTGTACACTGGAGTAACACAT	12655
QY	158	eIeuSerAlaAspAspIysThrVal-ArgLeuTrpAspHisAlaThrMetThrGluValI	178
Db	12654	TCTTTCTTGTGATGACAAAACGTTTAAGACTTTGGGATCATGCTACTATGACAGAAATGA	12595
QY	178	ySerLeuAsnPheAsnMetSerValSerSerMetGluTyrIleProGluGluIleI	198
Db	12594	AATCTCTAAATTTTATATGTTGCTCAGAGATGGAATATATCTCGAGGAGAGATTT	12535
QY	198	eValIleThrTyrTrpIyrgSerIleAlaPheHisSerAlaValSerLeuGluProIleI	218
Db	12534	TGGTATTAACCTATGCGTGATGCTGCTGTTGACAGTACAGTAAGTTTGGCAACCAATTA	12475
QY	218	ySerPheGluValaProIaThrIleAsnSerAlaSer**HisProGluIysGluPheI	238
Db	12474	AATCTTTCAAGCTCTCTGCAAAACATTAACCTGTGAATCTCTTACACCTTGAGAAAGATTTG	12415
QY	238	eValAlaGlyIysIynAspPheIysLeuTyrLysTyrAspTrpAsnSerGlyGluGluI	258
Db	12414	TTGTTGCGAGGTGGTGAAGATGTTAAACT-TAAAGTATGATGATATATGTGGAGAGAAAT	12356
QY	258	eGluIserTyrLysGlyHisPheGlyProIleHisCysValaIaGpHiserProAspGlyG	278

	Db	12355	INAAATCCACAAAGGACCTCTGGTCCATTACAGTGTGAGATTGATTCCTGATGGAG	12296
Qy	278	luLeuYrAlaSerGlySerGluAspGlyThrLeuArgLeuTrpGluThrValaGlyL	298	
Db	12295	AACTCATGTCCAGTGGTGTTCGAAAGATGGACAACCTGAGAC--TGGCAAACTGTGGTAGGAA	12238	
Qy	298	ysThrTygIleuTrpIlyCysVal**ProGluGluAspSerGlyGluLeuAlaLysP	318	
Db	12237	AAACATATGAGCCCTTGTGAAATGATGTGCTTCCTGGAAGAAATGCTGTGAGCTGGCAAAAC	12178	
Qy	318	rolYstIleGlyPheProGluThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluA	338	
Db	12177	CAAAAGATTGGTTTCCAAAGACAAAGAGAGAG-----CTAGAAAGAAATTCCTCAGAGA	12124	
Qy	338	snserAspSerIleTyrSerSerThrProGluValIlyAla	351	
Db	12123	ATTGAGATTCCATCTATTTCTTCTCAACTCCTCAAGTTAAGGCC	12083	
RESULT 8				
AAK78662/C				
ID	AAK78662	standard; DNA; 15332 BP.		
XX				
AC	AAK78662;			
XX				
DT	07-NOV-2001	(first entry)		
XX				
XX		Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33474.		
XX				
KM	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;			
XX	cytostatic; gene therapy; vaccine; metastasis; ds.			
OS	Homo sapiens.			
XX				
PN	WO200157182-A2.			
XX				
PD	09-AUG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US01354.			
XX				
PR	31-JAN-2000; 2000US-0179065.			
PR	04-FEB-2000; 2000US-0180628.			
PR	24-FEB-2000; 2000US-0184664.			
PR	02-MAR-2000; 2000US-0186350.			
PR	16-MAR-2000; 2000US-0189874.			
PR	17-MAR-2000; 2000US-0190076.			
PR	18-APR-2000; 2000US-0198122.			
PR	19-MAY-2000; 2000US-0205515.			
PR	07-JUN-2000; 2000US-0209467.			
PR	28-JUN-2000; 2000US-0214886.			
PR	30-JUN-2000; 2000US-0215135.			
PR	07-JUL-2000; 2000US-0216647.			
PR	07-JUL-2000; 2000US-0216880.			
PR	11-JUL-2000; 2000US-0217487.			
PR	11-JUL-2000; 2000US-0217496.			
PR	14-JUL-2000; 2000US-0218290.			
PR	26-JUL-2000; 2000US-0220963.			
PR	26-JUL-2000; 2000US-0220964.			
PR	14-AUG-2000; 2000US-0224518.			
PR	14-AUG-2000; 2000US-0224519.			
PR	14-AUG-2000; 2000US-0225213.			
PR	14-AUG-2000; 2000US-0225214.			
PR	14-AUG-2000; 2000US-0225266.			
PR	14-AUG-2000; 2000US-0225267.			
PR	14-AUG-2000; 2000US-0225268.			
PR	14-AUG-2000; 2000US-0225270.			
PR	14-AUG-2000; 2000US-0225447.			
PR	14-AUG-2000; 2000US-0225757.			
PR	14-AUG-2000; 2000US-0225758.			
PR	14-AUG-2000; 2000US-0225759.			
PR	18-AUG-2000; 2000US-0226279.			
PR	22-AUG-2000; 2000US-0226681.			
PR	22-AUG-2000; 2000US-0226868.			

PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228287.
 PR 01-SEP-2000; 2000US-0228343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0233968.
 PR 14-SEP-2000; 2000US-0233997.
 PR 14-SEP-2000; 2000US-0233998.
 PR 14-SEP-2000; 2000US-0233999.
 PR 14-SEP-2000; 2000US-0234000.
 PR 14-SEP-2000; 2000US-0234001.
 PR 14-SEP-2000; 2000US-0234063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 01-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 XX metastasis -
 PS Disclosure; SEQ ID NO 33474; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 15332 BP; 4490 A; 2686 C; 3180 G; 4976 T; 0 other;
 Alignment Scores:
 Pred. No.: 2, 11e-138
 Score: 1328.50 Length: 15332
 Percent Similarity: 66.59% Matches: 290
 Best Local Similarity: 63.74% Conservative: 13
 Query Match: 72, 20% Mismatches: 45
 DB: 22 Indels: 110
 Gaps: 3
 US-09-856-836-2 (1-351) x AAK78662 (1-15332)

Percent Similarity: 75.79%
 Best Local Similarity: 60.06%
 Query Match: 57.20%
 DB: 23 Gaps: 1

US-09-856-836-2 (1-351) x ABL06978 (1-1331)

```

OY 3 MetArgInThrProLeuThrcCysSerGlyHisThrArgProValValAsp***AlaPhe 22
DB 360 TTGGCAAAATCCCTGACCTGACAGTGGCCACAGCCAGCCGCTGCTCCAGCTGAC 419
OY 23 SerGlyIleThrProGlyIleThrPheLeuIleSerAlaCysLysAspGlyLysPromet 42
DB 420 AGCCACATCTGCGACGCGCTACTTCTCATCTGCGCTGCAAGAGATGGCAGCCGAG 479
OY 43 LeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaVal 62
DB 480 CTGCGGACGCGGACAGCGGCTGGCTGGCTGACCTTCGAGACACAAAGGCGCGTG 539
OY 63 TrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThr 82
DB 540 TGGAAAGCCAGCCTGACCGAATGCAACCTGCGCTCGGAGCGCGGCTGACCTGAC 599
OY 83 AlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLysHisIle 102
DB 600 GCGAAGTGTGGAAATCGGCTGACCGGCGCGGAGATACAGCTTCGACGACAAAGCAGATC 659
OY 103 ValLysThrValAspPheThrGlnAspSerAsnTrpLeu***ThrGlyGlyLysAspLys 122
DB 660 GGAAGAGGCTGCGCTGACAGGAGATTCGAGAACATGTCGTCGCGACGACAGAGAG 719
OY 123 LeuLeuArgIleTrpAspLeuAsnLysProGluAlaGluProLysGluIleSerGlyHis 142
DB 720 CTGGTGGCGGCTTCACTCACTGAGACGCGGAGCGGACGAGAGTACCGCGGACAT 779
OY 143 ThrSerGlyIleLysLysAlaLeuTrpCysSerAspLysGlnIleLeuSer---Ala 161
DB 780 ACGGGGCGCATACAGCGGCGCTCTTCTCCCGCGGACAAAGCTCATCTCGCGCGC 839
OY 162 AspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeuAsn 181
DB 840 CAGGACAAACAGCGCTGCGCTGCGGACCGGATGACGCGGATGAGGTGCGACCGCTCGCAG 899
OY 182 PheAsnMetSerValSerSerMetGluTrpIleProGluGlyLysLeuValIleThr 201
DB 900 TTCACAGCAATCCGACAGCTAGAGATCTCCAGCGACACACATCTGACCATATCG 959
OY 202 TyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLeyLysSerPheGlu 221
DB 960 CACGGCTCCCTGATTAAGCTTCTGCGAGATCGACACGCTGAAGAGCTGAAGAGGTGA 1019
OY 222 AlaProAlaThrIleAsnSerAlaSer***HisProGlyLysGluPheLeuValAlaGly 241
DB 1020 GTGCCACGACAGCTGCGCTGCGGACCTTTGCAATCCGATAGAGATGTGTTGTTGGC 1079
OY 242 GlyLysAspPheLysLeuTrpLysTrpAspTrpAsnSerGlyLysGluLeuGlnSerTrp 261
DB 1080 GCGGAGAGCTTAAAGATGTAAATTCGACTACATCACAGAAACGAAATTTGAATCTTC 1139
OY 262 LysGlyHisPheGlyProLeyHisCysValArgPheSerProAspLysGluLeuTrpAla 281
DB 1140 AAGGCAATTTGCGCGCTGACACAGCTGAAGTTGAGCCCGGATGCGGACACTGATGCC 1199
OY 282 SerGlySerGlyAspGlyThrLeuArgLeuTrpGlnThrValAlaGlyLysThrTrpGly 301
DB 1200 ACGGCTCGAGAGAGCGGCTGCGGCTGCGGACAGCAGCCGCTGGGCAAGACTTACGCG 1259
OY 302 LeuTrpLysCysVal***ProGluLysAspSerGlyLysGluLeuAlaLysProLys 319
DB 1260 TTGTGAAAGTGCACGAGCGCGGACCTGGCAATTCATCAGCTCGCGCGC 1313

```

RESULT 10
 ABL06978/c

```

ID ABL06978 standard; cDNA; 4891 BP.
XX
AC ABL06978;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 15416.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR MPI; 2001-656860/75.
XX
PS P-PSDB; ABB62875.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
Claim 1: SEQ ID NO 15416; 21pp + sequence listing; English.
XX
XX
The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB167737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at fltp.wipo.int/pub/publ/publshd.pct_sequences.
XX
XX
SQ Sequence 4891 BP; 1166 A; 1296 C; 1227 G; 1202 T; 0 other;
XX
XX
Alignment Scores:
Pred. No.: 4.31e-88 Length: 4891
Score: 876.50 Matches: 176
Percent Similarity: 63.99% Conservative: 55
Best Local Similarity: 48.75% Mismatches: 77
Query Match: 47.64% Indels: 55
DB: 23 Gaps: 4
XX
XX
US-09-856-836-2 (1-351) x ABL06978 (1-4891)
OY 38 AspGlyLysProMetLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGly 57
DB 1929 GATGGCAAGCCCGATGCTGCGGACAGCGGACCGGCTGGGCTGCGACCTTGCAGGGA 1870
OY 58 HisLysGlyAlaValAlaTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAla 77
DB 1869 CACAAGGCGCGCGCTGGAGAACGCCACGCTGAACCGAATGCGACCTGCGGCTCGGA 1810
OY 78 AlaAlaAspPheThrAlaLysValTrpAspAlaValSerGlyLysGluLeuMetThrLeu 97
DB 1809 GCGGCGGACTTCAACCGGCAAGGTGTGAATGCGTGAACCGGCGGCGGAGATACACAGCTTC 1750
OY 98 AlaHisLysHisIleValLysThrValAspPheThrGlnAspSerAsnTrpLeu***Thr 117

```



```

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```

```

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Alignment Scores:

```

Pred. No.: 6,54e-79
Score: 789.50
Length: 1753
Percent Similarity: 68.62%
Matches: 147
Best Local Similarity: 50.69%
Conservative: 52
Query Match: 42.91%
Mismatches: 88
DB: 21
Gaps: 2
Indels: 3

```

US-09-856-836-2 (1-351) x AAC43745 (1-1753)

```

QY 7 ProleuthrCysSerGlyHisThrArgProValValAsp***AlaPheSerGlyIleThr 26
DB 347 CCGCTGTGTCGACGACGCGCCTCGGCGCGGCTGCTTCTACGCCCCGTACA 406
QY 27 ProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLysPrometLeuArgGly 46
DB 407 CCGGATGGCTACTCTTCATCAGCGCCAGCAAGATCAATCCAAATGCTGTAATG 466
QY 47 AspThrGlyAspTyrPheGlyThrPheLeuGlyHisLysGlyAlaValTyrGlyAlaThr 66
DB 467 GAGACTGGCGACTGGATGGAGCTTTCAGGTCATTAAGAGAGCTGTGGAGCTGTGC 526
QY 67 LeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTyr 86
DB 527 CTTGACAGGAATGCTGTGAGGCGTCAATCTCTCTCTGCTTCTTACGTAAGATAGG 586
QY 87 AspAlaValSerGlyAspGlyLeuMetThrLeuAlaHisLysHisIleValLysThrVal 106
DB 587 GATGCACTTACTGGGAGTACCTTATTCATTTGACACCAAGCATATATAGTGGCGCTGT 646
QY 107 AspPheThrGlnAspSerAsnTyrLeu***ThrGlyLysGlnAspLysLysLeuArgIle 126
DB 647 GCCTTCTCTGAGATACACACCTTTTACTCACTGAGGACATGGAAGATTTGCGTGTG 706
QY 127 TyrAspLeuAsnLysProGlnAlaGlnProLysGlnLysSerGlyHisThrSerGlyIle 146
DB 707 TATGATATGAAATCGTCTGATAGCTCAAGAGAGCTTGCACAAATACCTGATCCGCT 766
QY 147 LysLysAlaLeuTyrCysSerAspAspLysGlnIleLeuSerAla-----AspAspLys 164
DB 147 LysLysAlaLeuTyrCysSerAspAspLysGlnIleLeuSerAla-----AspAspLys 164

```

```
Db 767 CGAAGCTGCTGCTGCTTCTAGTACCAACTATATAGTCTCTGCACTGACATGGGT 826
Oy 165 ThValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeuAsnPheAsnMet 184
Db 827 CGAGTAAAGGCTGGGATGTGACAACTGCAAAAATGTCCAAACTCTTGAACCAAGGCT 886
Oy 185 SerValSerSerMetGluTrpLeuProGluGlyGluLeuValIleThrTrpGlyArg 204
Db 887 TCTGTCACTAGTACGACGAACCAAGATGCGATTCATCTACTGCTGATGGCTCA 946
Oy 205 SerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGluAlaProAla 224
Db 947 AGGTGATAAATTCGGGACCGGATTCATCTTGGCTTGTAAACCTATGATATCCATGCG 1006
Oy 225 ThrIleAsnSerAlaSer***HisProGluLys---GluPheLeuValAlaGlyGlu 243
Db 1007 AAGTGGCAATCGGCTGCTCACTGACCTAAGCTGCGAGTAAATTTGCGCTGGAGAGAA 1066
Oy 244 AspPheLysLeuTrpLysTrpAspTrpAsnSerGlyGluLeuGluSerTrpLysGly 263
Db 1067 GATTTATGGGTTCATGATTTGATTTCTTCACGTGTAAGAAATAGCTTGAACAAAGGA 1126
Oy 264 HisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeuTrpAlaSerGly 283
Db 1127 CATCATGTGTCACGTCACACTGTGCTCGGTTTGCACTTGTGTGATCATATACGATCGGGT 1186
Oy 284 SerGluAspGlyThrLeuArgLeuTrpGln 293
Db 1187 TCAGAGATGGCACCATTGCGATCTGGCAG 1216

RESULT 13
AAC47397
ID AAC47397 standard; DNA: 1410 BP.
XX AAC47397;
AC 18-OCT-2000 (first entry)
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 53661.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX OS
XX PN EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 18-JUN-1999; 99US-0139817.
XX 21-JUN-1999; 99US-0139899.
XX 22-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140699.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
```

PR 23-JUL-1999; 99US-0145145-
 PR 23-JUL-1999; 99US-0145218-
 PR 23-JUL-1999; 99US-0145222-
 PR 26-JUL-1999; 99US-0145276-
 PR 27-JUL-1999; 99US-0145913-
 PR 27-JUL-1999; 99US-0145918-
 PR 27-JUL-1999; 99US-0145919-
 PR 28-JUL-1999; 99US-0145951-
 PR 02-AUG-1999; 99US-0146388-
 PR 02-AUG-1999; 99US-0146389-
 PR 03-AUG-1999; 99US-0146389-
 PR 04-AUG-1999; 99US-0147204-
 PR 04-AUG-1999; 99US-0147302-
 PR 05-AUG-1999; 99US-0147192-
 PR 05-AUG-1999; 99US-0147260-
 PR 06-AUG-1999; 99US-0147303-
 PR 06-AUG-1999; 99US-0147416-
 PR 09-AUG-1999; 99US-0147493-
 PR 09-AUG-1999; 99US-0147935-
 PR 10-AUG-1999; 99US-0148171-
 PR 11-AUG-1999; 99US-0148319-
 PR 12-AUG-1999; 99US-0148341-
 PR 13-AUG-1999; 99US-0148565-
 PR 13-AUG-1999; 99US-0148684-
 PR 16-AUG-1999; 99US-0149368-
 PR 17-AUG-1999; 99US-0149175-
 PR 18-AUG-1999; 99US-0149426-
 PR 20-AUG-1999; 99US-0149722-
 PR 20-AUG-1999; 99US-0149723-
 PR 23-AUG-1999; 99US-0149930-
 PR 23-AUG-1999; 99US-0149930-
 PR 25-AUG-1999; 99US-0150566-
 PR 26-AUG-1999; 99US-0150884-
 PR 27-AUG-1999; 99US-0151065-
 PR 27-AUG-1999; 99US-0151066-
 PR 30-AUG-1999; 99US-0151080-
 PR 31-AUG-1999; 99US-0151303-
 PR 01-SEP-1999; 99US-0151348-
 PR 01-SEP-1999; 99US-0151930-
 PR 07-SEP-1999; 99US-0152363-
 PR 13-SEP-1999; 99US-0153070-
 PR 15-SEP-1999; 99US-0153758-
 PR 15-SEP-1999; 99US-0154018-
 PR 16-SEP-1999; 99US-0154039-
 PR 20-SEP-1999; 99US-0154779-
 PR 22-SEP-1999; 99US-0155139-
 PR 23-SEP-1999; 99US-0155486-
 PR 24-SEP-1999; 99US-0155659-
 PR 28-SEP-1999; 99US-0156458-
 PR 29-SEP-1999; 99US-0156596-
 PR 04-OCT-1999; 99US-0157117-
 PR 05-OCT-1999; 99US-0157753-
 PR 06-OCT-1999; 99US-0157865-
 PR 07-OCT-1999; 99US-0158029-
 PR 08-OCT-1999; 99US-0158232-
 PR 12-OCT-1999; 99US-0158369-
 PR 13-OCT-1999; 99US-0159293-
 PR 13-OCT-1999; 99US-0159294-
 PR 13-OCT-1999; 99US-0159295-
 PR 14-OCT-1999; 99US-0159329-
 PR 14-OCT-1999; 99US-0159330-
 PR 14-OCT-1999; 99US-0159331-
 PR 14-OCT-1999; 99US-0159637-
 PR 18-OCT-1999; 99US-0159638-
 PR 21-OCT-1999; 99US-0160741-
 PR 21-OCT-1999; 99US-0160767-
 PR 21-OCT-1999; 99US-0160768-
 PR 21-OCT-1999; 99US-0160770-
 PR 21-OCT-1999; 99US-0160814-
 PR 21-OCT-1999; 99US-0160815-

PR 22-OCT-1999; 99US-0160980-
 PR 22-OCT-1999; 99US-0160981-
 PR 22-OCT-1999; 99US-0160989-
 PR 25-OCT-1999; 99US-0161404-
 PR 25-OCT-1999; 99US-0161405-
 PR 25-OCT-1999; 99US-0161406-
 PR 26-OCT-1999; 99US-0161359-
 PR 26-OCT-1999; 99US-0161360-
 PR 26-OCT-1999; 99US-0161361-
 PR 28-OCT-1999; 99US-0161920-
 PR 28-OCT-1999; 99US-0161992-
 PR 28-OCT-1999; 99US-0161993-
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
 Pred. 3.82e-78
 Score: 781.50
 Percent Similarity: 66.028
 Best Local Similarity: 47.908
 Query Match: 42.478
 DB: 21
 Gaps: 4

US-09-856-836-2 (1-351) x AAC47397 (1-1410)

QY 7 ProLeuThrcysSerGlyHisThrArgProValValAsp**AlaPheSerGlyIleThr 26
 DB 258 CCGTTAGTGTGCCAGGAGCATTCCTGCCGTCGATGCTTATATATCCGGTACT 317

QY 27 ProTyrGlyTyrPheLeuIleSerAlaCysLeuAspGlyIleThrProMetLeuArgGly 46
 DB 318 CCAAGTGGGTTCTTCTCATTTAGCGCCAGCAAGATTCGAATCCGATGTTGAGGAATGGA 377

QY 47 AspThrGlyAspTrpIleGlyThrPheLeuGlyHisLeuGlyAlaValTrpGlyAlaThr 66
 DB 378 GAGACTGTGACGATGGATTTGGACATTTGAAGGACATTAAGGACACTTGGATTCGACAC 437

QY 67 LeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTrp 86
 DB 438 CTTCGATTAATAATGCTATTCCTGCTGCCACGTTCTGATTTTCACCTCGAATAATATG 497

QY 87 AspAlaValSerGlyAspGlyLeuMetThrLeuAlaHisLysHisIleValLysThrVal 106
 DB 498 AATGCATTGACAGAGATGATTCACCTCTTGAACACAGACATTTGCTGCATGT 557

QY 107 AspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyIleLysLysLeuLeuArgIle 126
 DB 558 GCCTTTCTGAGGACACTCCCTTACTACACGTCGTCGGAATGAGAAATACCTTCGATA 617

QY 127 TyrAspLeuAsnLysProGluAlaGluProLysGluIleSerGlyHisThrSerGlyIle 146
 DB 618 TTGCAATTTGAATCGCCACGACCTCCAAAGAGTGGGAATTCCTCGTGCATTT 677

QY 147 LysLysAlaLeuTrpCysSerAspLysGlnIleLeuSerAla-----AspAspLys 164
 DB 678 AGAAGTGTCAATGCTTCTCATTAAGTATTAACATTTAAGCTCTTGCACACATACCGGT 737

QY 165 ThrValArgLeuTrpAspHisAlaThrMetThrGlyValLysSerLeuAsnPheAsnMet 184
 DB 738 GACATTAGGTTATGGACAAAGACATGTCATTAACATTTAAGCTCTTGCACACATACCGGT 797

QY 185 SerValSerSerMetGluTyrIleProGluGlyIleIleValIleThrTyrGlyArg 204
 DB 798 CCAATTACTAGTCTGCAAGTAACTGACGCGCATTAACACTGCTGATGATCT 857

QY 205 SerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGluAlaProAla 224
 DB 858 AGTGTTAGTTGGACGCTAAATAATTTGGATTCCTGAAGGCTATACATGCTTGC 917

QY 225 ThrIleAsnSerAlaSer**HisProGluLys--GluPheLeuValAlaGlyGly 243
 DB 918 AATGTTGAATCGGACATCGTTGGAACCGAACAACGTTAACACTTTGATTCCTGAGAGAGA 977

QY 244 AspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSerTyrLysGly 263

```
Db 978 GATATGTCGGGACACAGGTTGATTTCCAGACTGCAGAGAAAGATTGGTGAACAAGGT 1037
Oy 264 HisphecglyProIleHisCysValArgPheSerProAspGlyLeuTyrAlaSerGly 283
Db 1038 CATCACGGGCGACGTCACCTGCTGAGGTATGCGCCAGAGAGGGAGTCATACCTCAGGC 1097
Oy 284 SerGluAspGlyThrLeuArgLeuTyrPclnThrValValGlyLysThrTyrGlyLeuTyr 303
Db 1098 TCAGAGACGACGACGTCAGATATCG-----GTGGTGGTTCGCTAAC----- 1142
Oy 304 LysCysVal**ProGluGluAspSer 312
Db 1143 -----CATCCTGAGAGAGACAT 1160

RESULT 14
AC47955
ID AC47955 standard; DNA; 1540 BP.
XX
AC AC47955;
XX
DE 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55727.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128214.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
```

```
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
```



```

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160881.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Alignment Scores:

```

Pred. No.: 6,65e-77 Length: 1540
Score: 771.00 Matches: 156
Percent Similarity: 64.1% Conservative: 53
Best Local Similarity: 47.85% Mismatches: 103
Query Match: 41,908 Indels: 14
DB: 21 Gaps: 4

US-09-856-836-2 (1-351) x AAC47955 (1-1540)

QY 6 ThrProLeuThrcysSerGlyHisThrArgProValAlaSp***AlaPheSerGlyIle 25
DB ACCTCCGCTAGTGTCCATGCGCATTCGACACCTGTTGATTTGTTTCCACAGTCCAAATC 439
QY 26 ThrProGlyGlyThrPheLeuIleSerAlaCysLeuAspGlyLysProMetLeuArgGln 45
DB ACCTCCGATGTTCTTCCATCAGCGGAGAAAGATTCACCAATGTTGGAAAT 499
QY 46 GlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisGlyValAlaValTrpGlyAla 65
DB GCGCAAACTGGAGATTGGATTGGTACATTTGAAGTCATTAAGTCTGTCGAGTTCT 559
QY 66 ThrLeuAsnLysAspAlaThrLysAlaIleAlaAlaAlaAspPheThrAlaLysVal 85
DB TGTCTTGATTAACATGCTTTACGTCGACCTGCTGATTCGATTTTCACGAAAGCTT 619
QY 86 TrpAspAlaValSerGlyAspGlyLysGlyLeuMetThrLeuAlaHisLysHisIleValLysThr 105
DB TGGGATGCTTTGACCTGGGATGTCATTCCTTTGACACACACATTTTTCGAGCA 679
QY 106 ValAspPheThrGlnAspSerAsnTrpLeu***ThrGlyGlyLysAspLysLeuArg 125
DB TGGCCTCTCTCAGCATGCAAAATATCTAATCACAGGAGATTGCAAAATTTCTTCT 739
QY 126 IleTrpAspLeuAsnLysProGlnIleGluProLysGlnIleSerGlyHisThrSerGly 145
DB GTTTTGACTTGATTCGCTTGGATGACCTCCTACAGAAATGATTAATCTCTGTTCT 799
QY 146 IleLysLysAlaLeuTrpCysSerAspLysGlnIleLeuSerAla-----AspAsp 163
DB ATCAGAACACTAATACATGCTTCAGCGTATGACACATTAATTTTCTTGCATATAT 859
QY 164 LysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeuAsnPheAsn 183
DB GGTGGTGTGAGGTATGGAGTGTGAGAGTGGCAAAATATGCAAACTGTAACCAAG 919
QY 184 MetSerValSerSerMetGlyTrpIleProGlnGlyGlnIleLeuValIleThrTrpGly 203
DB TCTCCGTGTCACCGATGCTGAACTGAGTCAAGTGAAGCGATATTAACAATGCCATGG 979
QY 204 ArgSerIleAlaPheHisSerAlaValSerLeuGlnProIleLysSerPheGlnAlaPro 223
DB TCACCTGTTAAATTTTGGGATGCAATATTTTGGACAGTAAAGTTTACGACATGCC 1039
QY 224 AlaThrIleAsnSerAlaSer***HisProGlnLys---GluPheLeuValAlaGly 242
DB TCCAATATCGAATTCGACGCTTGAGCCAAATATGCGCAAAATTTGTTCTGTGGA 1099
QY 243 GluAspPheLysLeuTrpLysTrpAspTrpLysAsnSerGlyGlnGluLeuGlnSerTrpLys 262
DB GAAGATATGCTGGCTTCGACTCTTTCATTTTCCACACCGAAGAGGATTTGATGCAACAG 1159
QY 263 GlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGlnLeuTrpAlaSer 282
DB GACATCATGTCGTCGCTGCTGACGTGTGAGATTTGACCAACAGGATGATCTTAAGCTTCA 1219
QY 283 GlySerGluAspGlyThrLeuArgLeuTrpGlnThrValValGlyLysThrTrpGlyLeu 302
DB GGTCTGGAAGATGTACATTAAGATTTTGCAAACT----- 1255
QY 303 TrpLysCysVal**ProGlnGluAspSerGlyGlnLeuAlaLysProLysIleGlyPhe 322
DB 1256 ---GGACCGGTGAATTCCTGAGAGATTCAGCGAG-----TCAAAGCCGACCAACAGTGTG 1306

```

OY 323 ProglutThralagluu 328
||| |||
Db 1307 GATGAGCTTCCTCCTAG 1324
RESULT 15
AAC46123
ID AAC46123 standard; DNA; 1538 BP.
XX AAC46123;
AC
AC 18-OCT-2000 (first entry)
DT
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 48986.
DE
DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
OS
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123588.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR	17-AUG-1999;	99US-0149175.	
PR	18-AUG-1999;	99US-0149426.	
PR	20-AUG-1999;	99US-0149722.	
PR	20-AUG-1999;	99US-0149723.	
PR	20-AUG-1999;	99US-0149929.	
PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151065.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151348.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	23-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156590.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0158293.	
PR	13-OCT-1999;	99US-0159294.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Alignment Scores:			
Pred. No.:			8.61e-77
Score:			770.00
Percent Similarity:			64.1%
Best Local Similarity:			47.85%
Query Match:			41.85%
DB:			21
			1538
			156
			53
			103
			14
			4

```

Db      378  ACTCGGCTAGTGTGGCAATGGCCATTCGAGAACTGTTGGATTTGTGTCTACAGTCCAAATC 437
Oy      26  ThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLysProMetLeuArgGln 45
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      438  ACTCCGATGATGGTTTCTTCTCTCATCGCCGACAACTAAAGATTCACCAATGTTGGAAAT 497
Oy      46  GlyAspThrGlyAspThrIleGlyThrPheLeuGlyHisLysGlyValAlaValIleArgAla 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      498  GGGGAAACGCGAGATTTGGATTGGTACATTTGAAAGGTCATAAAGCGTCGTGTGGAGTTCT 557
Oy      66  ThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAlaAspPheThrAlaLysVal 85
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      558  TGTCTGTGAATACAAATGCTTACGTGACGCTTGCTGCTGATTTCTTTCACGGAAGCTT 617
Oy      86  ThrAspAlaValSerGlyLysProIleuMetThrLeuAlaHisLysHisIleValLysThr 105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      618  TGGGATGCTTGGATCGGGGATGCTTGGCATCTCTTTTGACACACACGATATTGTTCCAGCA 677
Oy      106  ValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyLysAspLysLeuArg 125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      678  TGGCCTTCTCTCAGATACGAAATATCTAATCAGAGGAGATTTGAGAAATATCTTCGT 737
Oy      126  IleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSerGlyHisThrSerGly 145
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      738  GTTTTGCACTGTAATGCTGTGATGACCCCTCAGCAAAATGATATAATCTCCTGTTCT 797
Oy      146  IleLysLysAlaLeuTyrPcysSerAspLysGlnIleLeuSerAla-----AspAsp 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      798  ATAGAACACTAACATGCGCTTCCACGGGATCAAAACAATATAAGTTCTTCCACTGATATT 857
Oy      164  LysThrValArgLeuThrAspHisAlaThrMetThrGluValLysSerLeuAsnPheAsn 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      858  GGTGGTGCTGAGTTATGGATGTGAGAGAGCGCAAAATAGTGAAACTCTGAGAACCCAG 917
Oy      184  MetSerValSerSerMetGlyTyrIleProGluGlyIleLeuValIleThrTyrGly 203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      918  TCTCCTCTACCGAGTGTGTGAAGTGAGTCAAGATGAGACGGTATATTAACAATGCCGATGG 977
Oy      204  ArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGluAlaPro 223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      978  TCAACTGTAAATTTGGATGTCGAAATCAATTTGGATGTAGTGAAGATTCAGCATGCCCC 1037
Oy      224  AlaThrIleAsnSerAlaSer**HisProGluLys---GluPheLeuValAlaGlyGly 242
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1038  TGCATATTCGAAATCTGCATGCGTTGACCCAAAATCTGGCAACAATATCGTTGGTGGA 1097
Oy      243  GluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSerTyrLys 262
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1098  GAAATATATGGGGTTCGACTCTTGATTTCCACACCGGAAAGAGATTTGATGTCACAAAG 1157
Oy      263  GlyHisPheGlyProIleHisCysValaIlePheSerProAspGlyGluLeuTyrAlaSer 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1158  GGAATCATATGCTCCGGTTCACTGTGTGAGATTTGCACACACAGGTGAGTCTTAAGCTTCA 1217
Oy      283  GlySerGlyAspGlyThrLeuArgLeuThrIleProIleHisValAlaGlyLysThrTyrGlyLeu 302
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1218  GGGTCTGAGATGGTACATACGATTTGGCGAAACT----- 1253
Oy      303  TrpLysCysVal***ProGluGluAspSerGlyLeuLeuAlaLysProLysIleGlyPhe 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1254  ---GGACCGGAGAAATCCGAAAGAGATCAGAG-----TCAAAGCCGAAAGCAGAGTGTG 1304
Oy      323  ProGluThrAlaGluGln 328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1305  GATGAGGTTGCTCGTAAAG 1322

Search completed: February 6, 2003, 13:11:41
Job time : 300 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2003, 10:00:47 ; Search time 71 Seconds

(without alignments)
7040.607 Million cell updates/sec

Title: US-09-856-836-1

Perfect score: 1630

Sequence: 1 ttaccgcctccgtagggga.....atgcctctaataaaaaa 1630

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5H_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5H_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PTUTS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1011.4	62.0	1820	US-08-828-922-2	Sequence 2, Appl1
2	240.8	14.8	686	US-09-040-984-36	Sequence 36, Appl1
3	240.8	14.8	686	US-09-123-912-36	Sequence 36, Appl1
4	240.8	14.8	686	US-09-643-597-36	Sequence 36, Appl1
5	50.6	3.1	7218	US-08-232-463-14	Sequence 14, Appl1
6	48.2	3.0	577	US-09-385-982-449	Sequence 449, Appl1
7	48.2	3.0	647	US-08-188-582-17	Sequence 17, Appl1
8	45.8	2.8	2152	US-08-646-715-17	Sequence 17, Appl1
9	45.8	2.8	2152	US-08-646-715-17	Sequence 17, Appl1
10	44.2	2.7	1221	US-09-489-506-2	Sequence 2, Appl1
11	44.2	2.7	1221	US-09-489-506-2	Sequence 2, Appl1
12	39.6	2.4	2085	US-08-283-917-8	Sequence 8, Appl1
13	39.6	2.4	2085	US-08-283-917-8	Sequence 8, Appl1
14	37.8	2.3	1607	US-08-883-534-2	Sequence 2, Appl1
15	37.8	2.3	1607	US-08-883-534-2	Sequence 2, Appl1
16	37.6	2.3	7042	US-09-204-764-2	Sequence 2, Appl1
17	37.6	2.3	7042	US-09-204-764-2	Sequence 2, Appl1
18	37.6	2.3	7042	US-09-435-115-1	Sequence 1, Appl1
19	37.6	2.3	7042	US-09-435-115-1	Sequence 1, Appl1
20	37.6	2.3	7042	US-09-690-364-21	Sequence 21, Appl1
21	37.6	2.3	7042	US-09-690-364-21	Sequence 21, Appl1
22	37.6	2.3	7042	US-09-690-364-21	Sequence 21, Appl1
23	37.6	2.3	7042	US-09-690-364-21	Sequence 21, Appl1
24	36.8	2.3	7881	US-08-751-189-1	Sequence 1, Appl1
25	36.8	2.3	7881	US-08-751-189-1	Sequence 1, Appl1
26	35.8	2.2	1368	US-09-060-836-1	Sequence 1, Appl1
27	35.6	2.2	1254	US-08-972-927-8	Sequence 8, Appl1
				US-09-134-001C-973	Sequence 973, Appl1

28	35.4	2.2	8743	US-09-081-320-1	Sequence 1, Appl1
29	35.4	2.2	8743	US-09-574-141A-1	Sequence 1, Appl1
30	35.4	2.2	8743	US-09-707-780-1	Sequence 1, Appl1
31	35.2	2.2	2343	US-09-641-638-652	Sequence 652, Appl1
32	35.2	2.2	4258	US-07-765-830A-5	Sequence 5, Appl1
33	35.2	2.2	2272	US-09-108-857-1	Sequence 1, Appl1
34	35.2	2.2	2272	US-09-108-857-1	Sequence 1, Appl1
35	34.6	2.1	2418	US-09-285-379-1	Sequence 1, Appl1
36	34.4	2.1	2487	US-08-377-292-1	Sequence 1, Appl1
37	34.4	2.1	3747	US-09-690-364-17	Sequence 17, Appl1
38	33.4	2.0	5152	US-09-690-364-17	Sequence 17, Appl1
39	33.3	2.0	4880	US-08-961-527-147	Sequence 147, Appl1
40	32.8	2.0	1119	US-09-402-929-5	Sequence 5, Appl1
41	32.8	2.0	1161	US-08-987-943-2	Sequence 2, Appl1
42	32.8	2.0	1755	US-09-134-001C-2770	Sequence 2770, Appl1
43	32.8	2.0	2520	US-08-987-943-1	Sequence 1, Appl1
44	32.8	2.0	3465	US-08-914-999-5	Sequence 5, Appl1
45	32.8	2.0	14176	US-08-307-499-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-828-922-2
Sequence 2, Application US/08828922
Patent No. 5834240
GENERAL INFORMATION:
APPLICANT: Olga, Bandman
APPLICANT: Pirelli, Lal
TITLE OF INVENTION: TRANSFORMING GROWTH FACTOR-B RECEPTOR
TITLE OF INVENTION: ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828, 922
FILING DATE: Herewith
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0258 US
TELEPHONE/COMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1820 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRATTU01
CLONE: 753423
US-08-828-922-2
Query Match 62.0%; Score 1011.4; DB 2; Length 1820;
Best Local Similarity 83.3%; Pred. No. 4.7e-299;

Matches 1312; Conservative 0; Mismatches 225; Indels 38; Gaps 13

OY	81	CCGCATATGGCATATGAGCAGACGCGCGCTACCTCTCGTGGGCAACAGCGCCGCTGTGG	140
Db	254	CCGCATATGGAATATGACAGACGCGCTACCTGTTCTGGACACAGCCACCGCTGTGG	313
OY	141	ATATGGCGCTTACCGGCATATCAGCGCTTACCGCTACTTCTTCGATCAGCGCTTGCAAAGATG	200
Db	314	ATTGGCGCTTCAATGGCATATCAGCGCTTATGGGATATTTCTTAATCAGCGCTTGCAAAGATG	373
OY	201	GCATGCCCATGCTCCGCCACGGAGATATCAGAGACTGGATGTGAACATTTTGGGTCATA	260
Db	374	GTAATCCTATGCTACGCCACGGAGATATCAGAGACTGGATGTGAACATTTTGGGTCATA	433
OY	261	AAAGTGTGTTTGGGTGCACTTGTAAATAAAGATGCCACCAAAAGCTGCAGACAGCTG	320
Db	434	AAAGTGTGTTTGGGTGCAACACTGTAAATAAAGATGCCACCAAAAGCTGCAGACAGCTG	493
OY	321	CAGACTTCACAGCCCAAGATATGGGATGCGGTCTCAGAGAGATGAATTTGATACCTGGCTC	380
Db	494	CAGATTCACAGCCCAAGATGTGGGATGCTGTCTCAGAGAGATGAATTTGATACCTGGCTC	553
OY	381	ATATGCAATATGTCAAGACTGTGGATTTACACAGATATAGCAATTTACTGTTAACTGGGG	440
Db	554	ATTAACACATGTGCACAGCTGTGGATTTACCCAGATATAGTAATTTTGTAAACGGGG	613
OY	441	GACAGATATAACTGCTCCGCATATATGACTTGAACAAACTGAAACAGACACTTAAGGAA	500
Db	614	GACAGATATAACTGTACGATATATGACTTGAACAAACTGAAACAGACACTTAAGGAA	673
OY	501	TCAGTGGCCACACTTCCTGTATATTAATAAAGCCTGTGGTCCAGTACAGATTAACAGATCC	560
Db	674	TTAGTGGCCACACTTCCTGTATATTAATAAAGCTGTGGTCCAGTACAGATTAACAGATTC	733
OY	561	TTTCAGCGGATGATTAATAACTGTTCGCGCTCTGGGATCATGCCAATGACAGAGTGAAT	620
Db	734	TTTCGCTGATGACAAAACGTTCGCACTTTGGGATCATGCTACTATGACAGAGTGAAT	793
OY	621	CTCTGAATTTAATATGTCTGTATAGCAGATATGAGATATTTCTCTAAGAGAGATTTGG	680
Db	794	CTCTGAATTTAATATGTCTGTATAGTATGAGATATATTTCTCTAAGAGAGATTTGG	853
OY	681	TTATTTACTTATGAGAGATATATGCTTTTCATATGTCACACTTAAGTCTGAGCCATTTAAAT	740
Db	854	TTATTAATTTATGAGAGATATATGCTTTTCATATGTCACACTTAAGTTTGGACCAATTTAAAT	913
OY	741	CCTTTGAAGCTCTGCGACCATCAATTTCTGCTCTTTTATCTACAGAGAGAGATTTCTG	800
Db	914	CCTTTGAAGCTCTGCGAACCATCAATTTCTGCTCTTTTATCTCTAGAGAGAGATTTCTG	973
OY	801	TTGCGGGGTGGAAGACTTTTAACTGTACAGATATGATTAATAACAGTGGAGAGAGATTAG	860
Db	974	TTGCGGGGTGGAAGATTTTAACTTTATTAAGTATGATTAATAAGTGGAGAGAGATTAG	1033
OY	861	AATCTTACAAAGCTCACTTTGGTCCATCTCACTGTGTGAGATTCAGTCTCTGATGGGCAAC	920
Db	1034	AATCTTACAAAGGACACTTTGGTCCATCTCACTGTGTGAGATTTAGTCTCTGATGGGCAAC	1093
OY	921	TCTATGCCAGCGGTCTTAATATGGGACATTTGATTTGGGCAAACTGTGGTAGGAAGA	980
Db	1094	TCTATGCCAGTGTTCAGAAATATGAACATTTGAGACTATGGCAAACTGTGGTAGGAAGAA	1153
OY	981	CCATATGGCTGTGGAATATGCTGTTCCTGTAGAGAACACAGCGGGAACTGGCAAAAGCCAA	1040
Db	1154	CGATATGGCTTTGGAAATATGTGCTCTCTGAAGAAAGATATGTGTAGCTGGCAAAAGCCAA	1213
OY	1041	AGATGGCATTTTCCAAAAACGACAGAGAGAGACTGGCAGAAAGAAATTCCTTACAGAGATTT	1100
Db	1214	AGATGTGTTTCCAAAGACACAGAGAGAGAGCT--AGAGAAATTTGCTTACAGAGATTT	1270
OY	1101	CAATTTCCATATTTATCAATCAACTCTGTGAAGTTAAGGCTCAGAGATCAAGACGTGTCTGCT	1160
Db	1271	CAATTTCCATATTTCTTTCAGCTCTCTATGTTAAGGCTCAGAGCTCAATCATATGTGTGA	1330

QY	1161	GAAACCACT--ATGTCATGAGCTAAACAGACGAGACAGATCCGCTT--CAAGTTAC	1217
Db	1331	GTTGTTATACACTGACTATAAACACAGACGAGAAAAAGCATCAGCCTTCACAGATTAC	1390
QY	1218	TGTTGCTGAGGCGAAGAGGCGACGAAATATTTGGGCATATGATTAAGTCCAGTGCAC	1277
Db	1391	TGTCGTGCTTAAGGCGACGAACAGCAGTAAATATATGAGGAAAAATGAAATTAAGTCACAGTCTG	1450
QY	1278	GAAAGCT-----ACACAGTGTGGCCGAGAGGAAATATGGCTGAGTGTGAGAGTGCAG	1332
Db	1451	GAAACAATACTAACTTGTGGTTTACCTGTATAGTGAATA-----CTCAATGTGCATGTAAG	1506
QY	1333	GCAGAGAGA--TTGTGCTCACATAGTGCATAGCTGCTGCTTT-----GGAATGAAA	1382
Db	1507	GGAGGTGAGATTATCCCTTATAGTACAGAGGCGCTGTATCTTTTAATGAAATATATCA	1566
QY	1383	AGGCAACTTAATCTCCATT--TTACACCTAAATCTTTAGCTGTATATGT--TATG	1438
Db	1567	AGCCAACTCCCAATTTCTATTATTACAAATTAAGGCTTCTGTAGCTGTTAATGTTAAATATG	1626
QY	1439	AAGAGAAAAATATATTTGGCCATTTTTCTGACTTCCCTTAAAGAAATGCCCTTTTG	1498
Db	1627	GAGAGAAACCTATATTTGGCTGCATTTTTTCTGAT-----CTTAAACAGAAATGCCCTTTTCT	1682
QY	1499	TCCTTGGCT--AGTATGAGAGAGAGAGAAATACATGATTAATTAACCGGTGGATCTCT	1556
Db	1683	TTTTTTGCTTAGAGTTGTAAAGAGAGAGGAATATATATAAAGTAAGTAAGTGGTTTGATTTCC-	1740
QY	1557	TCATTGTACAAGAGCTGCTTCAGAACAGCTCA--TATTTTATGTTATCTAAATAAAAATGCC	1615
Db	1741	TGCTTCATTTAGTACCTGCTCTGAAACATCTAATTTGTTTATGTTGTCTAAATAAAAATGCC	1800
QY	1616	TCTAAATATAAAAAA 1630	
Db	1801	TCTAAACAAAAACAA 1815	

RESULT 2
 US-09-040-984-36
 ; Sequence 36, Application us/09040984
 ; Patent No. 6210883
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSTICS
 ; TITLE OF INVENTION: OF LUNG CANCER
 ; NUMBER OF SEQUENCES: 86
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and HERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSO for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/040,984
 ; FILING DATE: 18-MAR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.456
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-622-4900
 ; TELEFAX: 206-282-6031
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 36:


```

QY      837  ATTATTAACAGCGGGGAAGAGATTACAATCTTACAAGAGTCACTTTGGTCCATTCAGTGG  886
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      141  ATTGGCGCCTTGGAAAGAAAGATTGTGGAAGAGTCAAGGGCTCACTTTGGACCTATCAACAGTG  82

QY      897  TGAGATTCAGTCTCTGATGGGGAACCTATGCGCAGCGGTCTTGCGAAGATGG  945
      |  ||  |||  ||||  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      81  TTGGCTTCATCTGATGCGCAAGAGCTTACAGCAGCGGCGGCGGAGATGG  33

RESULT 8
US-08-188-582-17
: Sequence 17, Application US/08188582
: Patent No. 5534410

GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36

```

RESULT 8
 US-08-188-582-17
 Sequence 17, Application US/08188582
 Patent No. 5534410
 GENERAL INFORMATION:
 APPLICANT: Tjian, Robert
 APPLICANT: Comai, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoiko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
 TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

[illegible]

811 AGAGACCTTAAACTGTACACGATATGATTATAACAGTGCGAAGAGTTAGANATCCTACNA 870
| | | | |
1650 TGCACCAACAAAACCCGCCGGCCCCGCGTCGATTAAGAAAAACAAGCATGGAAGAAGGACGCT

OY	177	TTTCGATCAGCGCTTGCAAGAATGCGAACCCCATGCTCCGCCAGGAGATATAGGAGCT	236
Db	1208	TTTATGGCTCTCTGGCTTCAGAGAGATCGACATTAATTAAGGTGTGGATTATGACTGGAGATT	1267
OY	237	GGATTGGCAACATTTTGGGTCATTAAGAGTGCCTTTGGGGGTCCCAACATTTGAATTAAGGATG	296
Db	1268	TTTGACGACACCTCTTAAGGGGCATACGACTCTGTACAGATATTTTCATTCGACACACAGTG	1327
OY	297	CCACCAAAAGCTGGGACGACGAGCTCAGAGCTTCACAGCCCAAAGTATGGGATGCGGCTCAG	356
Db	1328	GCAGAGCTTTCTGGCTTCATGTCTTCCGAGATATGACCATTAAGCATATGGATTTTTCAGGGCT	1387
OY	357	GAGATGAATTTGATGACCCCTG---GCTCATAGACATTTGTCAAGAGCTGTGGATTTCAAC	413
Db	1388	TTTGATGCAATCAGAACCATCATCATGAGCCATCACCAATGTTCTTCTTACAGAGCCATCATCG	1444
OY	414	AGGATAGCAATTTACCTCCTTTAACTGGGGGACAGGATTAACAGCTGCGCATATATGACTTGA	473
Db	1448	CCAAATGGAGATCATATAGTGTCTGCTCAAGGATTAATAACTATTAATAATGTGGGAAGTGC	1507
OY	474	ACA	476
Db	1508	AAA	1510

RESULT 13
 US-08-961-716-8
 Sequence 8, Application US/08961716
 Patent No. 5880272
 GENERAL INFORMATION:
 APPLICANT: ADACHI, HIDEKI
 APPLICANT: TSUJIMOTO, MASAFUMI
 APPLICANT: INOUE, KEIZO
 APPLICANT: ARAI, HIROYUKI
 TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
 TITLE OF INVENTION: AND GENE THEREOF
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER &
 ADDRESSEE: NEUSTADT, P.C.
 STREET: 1755 S. Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,716
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/283,917
 FILING DATE: 03-AUG-1994
 APPLICATION NUMBER: JP 209943/1993
 FILING DATE: 03-AUG-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Obion, No. 5880272man F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 2292-030-0
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 413-3000
 TELEFAX: (703) 413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2085 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown

```

; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 844..2073
;
US-08-961-716-8

```

Query Match	2.4%;	Score 39.6;	DB 2;	Length 2085;
Best Local Similarity	47.9%;	Pred. No. 0.076;		
Matches 145;	Conservative	0;	Mismatches 155;	Indels 3;
				Gaps 1;

Qy	177	TTTCGATCAGGGCTTGC	AAAGTGGGCAAGGCCATGCTCCGCCAGGGAGATACAGGAGCT	236
Db	1208	TTATGCTCTCGCTTC	AGAGGATGCTACAAATTAAAGTGTGGGATTATGAGACTGGAGATT	1267
Qy	237	GGATTGGAACTTTTTTGG	GTCTATAAAGCTGTCTTTTGGGCTGCACATTTGAATTAAAGACT	296
Db	1268	TTTGACGCAACTCTTAA	GGGGCATACGACCTCTTACAGGATATTTTATTTCGACCAAGTG	1327
Qy	297	GCACCAAAAGCTGGC	CACAGCTGCAGACTTTCACAGCCAAAGTATGGATGGCGTCTCAG	356
Db	1328	GCAAGCTTCTGGCTT	CAATGTTCTTGCACATATTGACCAATTAAAGTATGGGATTTTCAGGGCT	1387
Qy	357	GAGATGAATTGATAC	CCCTG---GCTCATAGAACAATTGTCAAGACTGTGGGATTTTCACAC	413
Db	1388	TTTGATGTGATAGAA	CCATGCATGAGCCCATGCAACCAATGTGTTCTTAGAGGCATATATGC	1447
Qy	414	AGGATAGCAATTAC	CTGTAACATGGGGAGAGATAAACATGCTTGCACATATATGACTTGA	473
Db	1448	CCAAATGGAGATCAT	ATATAGTGTCTCTCAAGGGATTAACATATATAAAATATGTGGGAAGTGC	1507
Qy	474	ACA	476	
Db	1508	AAA	1510	

RESULT 14
 US-08-083-534-2
 Sequence 2, Application US/08883534
 Patent No. 5846777
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Lal, Preeti
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: TWO NEW WD-40 PROTEINMS
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/883,534
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0332 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166

QY	129	GGCCCGTGGTGGATTTWGGCTTCAGGGGACACAGCCCTACGGCTACTTTCGTATCAGCG	188
Db	450	GACATGTCTTTGATGATGTGATTTGTTGGCAGTTTTTCCATCAGGCTGTGTCTCTGAGTG	509
QY	189	CTTGGCAAGATGGCAGGCCATGCTCCGGCCAGGAGATACAGGAGCTGGATTTGGAACT	248
Db	510	GGGGAATGATGGATGCCAGCTGAAGATATGTGTACGTGTGAAGATGCTAGCTGTGGTGGTACT	569
QY	249	TTTTTGGGTCAATAAAGGTGTCTTTTGGGGGTGCACATTTGAATTAAGATGCCACCAAAAGCTG	308
Db	570	TCAAAGGTACACAAAGGAGGATCTCTGGATACAGGCATCTTTGATCGGGGAGGAGATGTGG	629
QY	309	CGACAGCAGCTGCGACTTTCACAGCCCAATATGTGGAT	346
Db	630	TGTCTGTCTTCGAGATGGGACACACACAGACTTTGGGAT	667

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2003, 10:00:47 ; Search time 2440 Seconds

(without alignments)
10819.116 Million cell updates/sec

Title: US-09-856-836-1

Sequence: 1 ttaccgcgcacgcgtaggga.....atgcctctaataaaaaa 1630

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estnu:*
6: em_estnu:*
7: em_estnu:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_hum:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1009.6	61.9	1860	11	AF161496 Homo sapi
2	872	53.5	933	14	BQ936632 AGNCOURT
3	819	50.2	948	13	B1407527 602918767
4	720.8	44.2	904	9	AU123736
5	703.8	43.2	752	12	BF023572
6	667	40.9	944	9	AL517952

7	658.8	40.4	684	9	AV060640
8	658	40.4	880	14	BO219033
9	651.8	40.0	702	12	BF720733
10	648.8	39.8	1174	14	BM802585
11	645.8	39.6	1058	12	BF682981
12	635.2	39.0	902	14	BQ889506
13	632.4	38.7	926	13	BI687709
14	631.6	38.7	875	14	BQ439153
15	631	38.7	1005	13	BI408095
16	629.8	38.6	1055	13	BM472016
17	629.4	38.6	839	12	BQ066448
18	624.6	38.3	656	13	BI112632
19	614.2	37.7	1019	13	BM450309
20	612.2	37.6	800	9	AU123939
21	611.6	37.5	901	9	AL518570
22	609	37.4	881	9	AU132581
23	608.2	37.3	757	12	BG687473
24	607.4	37.3	632	10	AM744196
25	605	37.1	620	12	BF467887
26	603.6	37.0	901	9	AL556515
27	599	36.7	1086	14	BQ433088
28	597.6	36.7	879	14	BQ222878
29	596.8	36.6	673	12	BF723058
30	588.6	36.1	1038	9	AL545602
31	585.2	35.9	894	12	BF118711
32	580.4	35.6	874	9	BQ883467
33	579.8	35.6	885	14	BQ672095
34	577.6	35.4	955	14	BO672095
35	576.4	35.4	833	9	AU119814
36	573	35.2	796	12	BG778828
37	571.6	35.1	1093	9	AL545569
38	570	35.0	959	9	AL547128
39	569.8	35.0	962	9	AL539156
40	569.4	34.9	704	12	BG740119
41	568.6	34.9	951	9	AL551950
42	563.4	34.6	629	10	AM682227
43	562.2	34.5	973	9	AL547220
44	561	34.4	664	13	BI596603
45	560	34.4	952	9	AL539237

ALIGNMENTS

RESULT 1	AF161496	1860 bp	mrna	linear	HTC 22-MAY-2001
LOCUS	AF161496				
DEFINITION	Homo sapiens HSPCL47 mRNA, complete cds.				
ACCESSION	AF161496				
VERSION	AF161496.1	GI:6841515			
KEYWORDS	HTC.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1860) Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G., Shen, Y., Fan, H.Y., Lu, G., Zhong, M., Xu, X.R., Han, Z.G., Zhang, W., Tao, J., Huang, Q.H., Zhou, J., Hu, G.X., Gu, J., Chen, S.J., and Chen, Z.				
TITLE	Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells				
JOURNAL	Genome Res. 10 (10), 1546-1560 (2000)				
MEDLINE	20499367				
PUBMED	11042152				
REFERENCE	2 (bases 1 to 1860) Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L., Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J., and Chen, Z.				
AUTHORS	Unpublished 3 (bases 1 to 1860) Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L., Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J., and Chen, Z.				

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 933)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .933

BASE COUNT	256 a	218 c	174 g	285 t
ORIGIN				

Query Match	53.5%	Score 872;	DB 14;	Length 933;
Best Local Similarity	98.2%;	Pred. No. 6.9e-226;		
Matches 913;	Conservative 0;	Mismatches 12;	Indels 5;	Gaps 3

703	16ATGTCATGAGCGATGAGTCTGTGGAGCATTAAATCCTTT--GAGGCTCTCGACACA	761
933	TCCTTTAAATAGTCAGATTAATCTGAAGCCATTAAATTCTTTGGAAAGCTCTCGACACA	874
QY	762 TCAA--TTTCGCGTCNNTTCATCCAGAAAGAGTTCTTGTGCGGGGTGAGAAAGACTTT	820
Db	873 TCAATTTGCGCTCTCTTATCCAAAGAAAGAGTTCTGTTGGCAGGGGAGAAAGACTTT	814
QY	821 AAACGTACAGATGATTATPACAGTCGAGAGAGATAGAAATCCTCMAAGATCATT	880
Db	813 AAACGTACAGATGATTATPACAGTCGAGAGAGATAGAAATCCTCMAAGATCATT	754
QY	881 GGTCCCATTCATGTGTGATGATTCAGTCTCTATGGGGAACCTATGCGCAGCGGTTCTGAA	940
Db	753 GGTCCTCATTCAGTGTGATGATTCAGTCTCTATGGGGAACCTATGCGCAGCGGTTCTGAA	694
QY	941 GATGGAGCATTCAGATTTGTGGCAACCTGGTAGGAAAGACCTATGGCCCTGTGGAATGC	1000
Db	693 GATGGAGCATTCAGATTTGTGGCAACCTGGTAGGAAAGACCTATGGCCCTGTGGAATGC	634
QY	1001 GTGATTTCTGAGAGACAGCGGGGAACTGCGCAAAAGCCAAAGATCGGATTTCCAGAAACA	1060
Db	633 GTGCTCTCTGAGAGACAGCGGGGAATCTGCCAAAGCCAAAGATCGGATTTCCAGAAACA	574
QY	1061 GAGAGGAAGAGCTGCGCAGAGAAATTTGCTTCAGAGAAATTCAGATTTCAATTCATCA	1120
Db	573 GCGAGGGAAGAGCTG--GAAAGAAATTTGCTTCAGAGAAATTCAGATTTCAATTCATCA	517
QY	1121 ACTCTGGAAGTAAAGGCTCTGAGATTCAGAGCTGTGCTGCCGGAAGCAATATGTCATAGGAC	1180
Db	516 ACTCTGGAAGTAAAGGCTCTGAGATTCAGAGCTGTGCTGCCGGAAGCAATATGTCATAGGAC	457

QY	1181	TAAGCAGCGAGACAAAGCATCGGCTTCAGACTACTGTCTGCTCTAGGCAAGAGGGC	1240
Db	456	TAAGCAGCGAGACAAAGCATCGGCTTCAGACTACTGTCTGCTCTAGGCAAGAGGGC	397
QY	1241	AGAAATATTGGGCGATATGAGTTAGCTCCAGTGCAGCAAGACACTACTCAGTGTGGCCG	1300
Db	396	AGAAATATTGGGCGATATGAGTTAGCTCCAGTGCAGCAAGACACTACTCAGTGTGGCCG	337
QY	1301	TGAGTCAAAATGGCTCAGTGTCTTGAGCTGCAGCGAGAGAGATTGTGTCACATAGTGCCA	1360
Db	336	TGAGTCAAAATGGCTCAGTGTCTTGAGCTGCAGCGAGAGAGATTGTGTCACATAGTGCCA	277
QY	1361	TAGCTGTGTTGGAGATGAAAAGCCAACTTACATCTCCATTTCACACTTAATTTCTT	1420
Db	276	TAGCTGTGTTGGAGATGAAAAGCCAACTTACATCTCCATTTCACACTTAATTTCTT	217
QY	1421	TTAGCTGTTTATGTTTATGAAAGAAAGAAATATATTTGGCCTATTTTCTGACTTCCCTTA	1480
Db	216	TTAGCTGTTTATGTTTATGAAAGAAAGAAATATATTTGGCCTATTTTCTGACTTCCCTTA	157
QY	1481	AAGAAGATATGCCCTTTTGTGCTTGGCTACAGTATGAGAGAGAGAAATACATGATTAAGTA	1540
Db	156	AAGAAGATATGCCCTTTTGTGCTTGGCTACAGTATGAGAGAGAGAAATACATGATTAAGTA	97
QY	1541	ACCGGTTTGAATCTCTTTCATTTGATACAGGACTGCTTCAGAACAGCTCATATTTTATGTA	1600
Db	96	ACCGGTTTGAATCTCTTTCATTTGATACAGGACTGCTTCAGAACAGCTCATATTTTATGTA	37
QY	1601	TCTAAATAAATGCCCTTAAATTAAGAAAAA	1630
Db	36	TCTAAATAAATGCCCTTAAATTAAGAAAAA	7

RESULT	3
BI407527/c	
LOCUS	948 bp mRNA linear EST_14-AUG-2001
DEFINITION	602918767r1 NCI-GAP_Lu33 Mus musculus cDNA clone IMAGE:5055736 3'
	mRNA sequence.
ACCESSION	BI407527
VERSION	BI407527.1 GI:15168450

```

FEATURES
Source
Location/Qualifiers
1. 948
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5055736"
/clone_idb="NCI CGAP Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10b (phage-resistant)"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: NotI; Site:2: EcoRI; 1st

```

strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - Oligo(dT) primer [5', TGTTACCAATCTCAGAGGAGGAGCGCCGCTCTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV33 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 261 a 237 c 176 g 274 t
ORIGIN

Query Match 50.2%; Score 819; DB 13; Length 948;
Best Local Similarity 97.0%; Pred. No. 1.9e-211;
Matches 919; Conservative 0; Mismatches 17; Indels 11; Gaps 8;

```

OY 633 ATATGCTGTAGCAGCATGGATGATATCTCTGAGAGAGATTTGGTTATTTACTTATG 692
Db 947 AATGTCCTTGAGCAGCATGGATGATATCTCTGAGAGAGATTTGGTTATTTACTTATG 888
OY 693 GAGCATCTATGCTTTTCATATGATGAGTAACT--CTGGAGCCATTTAACTCTTGAAGC 750
Db 887 GAGCATCTA-TGCTTTTCATATGATGAGTAACTTTCTGGAGCCATTTAACTCTTGAAGC 829
OY 751 TCTGCGACATCAATCTCTGCTC-TTTTCATCAGAGAGAGATTTCTTGTGCGGCTG 809
Db 828 TCTGCGACATCAATCTCTGCTCCTTTCATCCAGAGAGAGATTTCTTGTGCGGCTG 769
OY 810 GAGAGAC-TTTAACTGTACAGATGATATTAACATGAGAGATTTCTTGTGCGGCTG 867
Db 768 GAGAGACTTTTAACTGTACAGATGATATTAACATGAGAGATTTCTTGTGCGGCTG 709
OY 868 CAAAGTCACTTTGGTCCCATTCATCTGTGAGATTCAGTCTGATGGGAACTCTATGC 927
Db 708 CAAAGTCACTTTGGTCCCATTCATCTGTGAGATTCAGTCTGAT--GGAACTCTATGC 650
OY 928 CAGCGGTTCTGAAGTGGAGATTTGTTGGCAACTGTGTGAGAGAGAGAGAGAGAGAG 987
Db 649 CAGCGGTTCTGAAGTGGAGATTTGTTGGCAACTGTGTGAGAGAGAGAGAGAGAGAG 590
OY 988 CCGTGGAGAAATGGCGTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047
Db 589 CCGTGGAGAAATGGCGTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 530
OY 1048 ATTTCAGAAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1107
Db 529 ATTTCAGAAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473
OY 1108 CATCTATTCATCAATCTCTGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1167
Db 472 CATCTATTCATCAATCTCTGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 413
OY 1168 TATGTTCAATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1227
Db 412 TATGTTCAATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 353
OY 1228 AGGCAAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1287
Db 352 AGGCAAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 293
OY 1288 TGAAGTTCCTGAGTGAAGATGAGTGTCTGAGAGTGCAGAGAGAGAGAGAGAGAGAG 1347
Db 292 TGAAGTTCCTGAGTGAAGATGAGTGTCTGAGAGTGCAGAGAGAGAGAGAGAGAGAG 233
OY 1348 TCACATAGTGCATAGCCTGCTGTTGGAATGAAAGCCAACTTACAACTTCATTTTAC 1407
Db 232 TCACATAGTGCATAGCCTGCTGTTGGAATGAAAGCCAACTTACAACTTCATTTTAC 173
OY 1408 ACCATAATTTCTTTTACGTTGTTATTTATTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1467
Db 172 ACCATAATTTCTTTTACGTTGTTATTTATTTGAAGAGAGAGAGAGAGAGAGAGAGAG 113
OY 1468 TCACATTTCTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527

```

Db 112 TGACTTCCTTAAG 53
OY 1528 ACATGATTAAGTAACTACCGGTTGATCT-CTTTCATTTGATGAGAGAGAGAGAGAG 1573
Db 52 ACATGATTAAGTAACTACCGGTTGATCTCTTTCATTTGATGAGAGAGAGAGAGAGAG 6

RESULT 4
LOCUS AU123736 904 bp mRNA linear EST 01-AUG-2002
DEFINITION AU123736 NT2RM2 Homo sapiens cDNA clone NT2RM2000935 5', mRNA
sequence.

ACCESSION AU123736
VERSION AU123736.1 GI:10948452
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
Ota, T., Nakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
Nakamura, Y., Nishikawa, T., Negai, T., Suzuki, Y., Sugano, S. and
Isogai, T.

TITLE HRI human cDNA project (Ota, T., Nakamatsu, A., Ozawa, M., Ishii, S.,
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Negai, T., Suzuki,
Y., Sugano, S., Isogai, T.)
JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomisehri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers

FEATURES
source 1..904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM2000935"
/clone_id="NT2RM2"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

BASE COUNT 257 a 173 c 214 g 256 t 4 others

Query Match 44.2%; Score 720.8; DB 9; Length 904;
Best Local Similarity 89.3%; Pred. No. 9.7e-185;
Matches 806; Conservative 0; Mismatches 94; Indels 3; Gaps 3;

```

OY 64 CCGCCCTCCCTCCCTCCCTCCGCAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
Db 2 CCGCCCTCCCTCCCTCCCTCCGCAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
OY 124 CAGCGGCGCGGTGGTGGATTTGGCTTCAGCGGAGATCAGCGCTTACGCTTCTTGAAT 183
Db 62 CAGCGGCGCGGTGGTGGATTTGGCTTCAGCGGAGATCAGCGCTTACGCTTCTTGAAT 121
OY 184 CAGCGCTTGCAAAAGATGCGAAGCCCATGCTCCCGCAGGAGAGATACAGAGAGAGAGAG 243
Db 122 CAGCGCTTGCAAAAGATGCGAAGCCCATGCTACGCGCAGGAGAGATACAGAGAGAGAGAG 181
OY 244 AACATTTTGGGTCATAAAGTGTCTGTTGGGTGGAACATTGAATTAAGATGCCACCA 303
Db 182 AACATTTTGGGTCATAAAGTGTCTGTTGGGTGGAACATTGAATTAAGATGCCACCA 241
OY 304 AGCTGGACAGCAGCTGCAGACTTCAAGCCAAAGATGAGAGAGAGAGAGAGAGAGAGAG 363

```


DB 97 GATTAAGTAAACCGTTGATCTCTTTCATTTGACAGAGCTGTCGACAGCTCATAT 38
 OY 1592 TTTTATTTATCTAAATAAATGCTCTAAATATAAAAA 1628
 DB 37 TTTTATTTATCTAAATAAATGCTCTAAATATAAAAA 1

RESULT 6
 AL517952 944 bp mRNA linear EST 13-FEB-2001
 LOCUS AL517952
 DEFINITION AL517952 LITL_NFL011_NBC1 Homo sapiens CDNA clone CS0DA004YH09 5
 ACCESSION AL517952
 VERSION AL517952.1 GI:12781445
 KEYWORDS prime, mRNA sequence.
 SOURCE EST.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 944)
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length CDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..944
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DA004YH09"
 /clone_lib="LITL_NFL011_NBC1"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand CDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-stranded CDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifestech.com URL: http://fulllength.invitrogen.com"

BASE COUNT 259 a 197 c 235 g 250 t 3 others
 ORIGIN

Query Match 40.9%; Score 667; DB 9; Length 944;
 Best Local Similarity 89.9%; Pred. No. 4.3e-170;
 Matches 712; Conservative 1; Mismatches 79; Indels 0; Gaps 0;

OY 66 TCCCTCCCTCCCTGTCGCCATGCGCATGAGCGAGCGCGCTCACTGCTGGGCCACA 125
 DB 138 TCGGTGGCTTCCCGCCGCGCATGCGCAATGAGACGCGCGCTCACTGCTGGGCCACA 197
 OY 126 CGCGGCCCTGTGATNTGGCTTACAGGCGATCAGCGCTTACGGCTACTTCTGATCA 185
 DB 198 CGCGACCCCTGTGATNTGGCTTACAGGCGATCAGCGCTTACGGCTACTTCTGATCA 257
 OY 186 GGGCTTGCAGAGTGGCAGCCCATGCTCCGCGAGAGATACAGAGACTGATTTGAA 245
 DB 258 GCGCTTGCAGAGTGGTAAACCTATCTACGCGAGAGATACAGAGACTGATTTGAA 317
 OY 246 CATTTTGGGTATTAAGTGTGTTGGGGTCAACATTGAATTAAGATCCACCAAG 305
 DB 318 CATTTTGGGTATTAAGTGTGTTGGGGTCAACATTGAATTAAGATCCACCAAG 377
 OY 306 CTGCGACAGAGCTGACAGCTTACAGCCAAATATAGGATGGGCTCTCGAGATGAT 365

DB 378 CAGCTACAGAGCTGACGATTTTCACAGCCAAAGTGGAGTGTCTGTCAGAGATGAAT 437
 OY 366 TGATGACCCGCGGCATTAAGACATTTGCAAGACTGTGATTTTCACAGAGATGAAT 425
 DB 438 TGATGACCCGCGGCATTAAGACATTTGCAAGACTGTGATTTTCACAGAGATGAAT 497
 OY 426 ACTGNTAACTGGGGGACAGAGTAAACTGTCGCCATATATGACTTGAAACCACTGAG 485
 DB 498 ATTTTGTAAACGGGGGACAGAGATTAAGTGTACGATATATGACTTGAAACCACTGAG 557
 OY 486 CAGAACTTAAGGAATTCAGTGGCCACACTTCTGATTTAAAGGCTGTGGTGGAGT 545
 DB 558 CAGAACTTAAGGAATTTAGTGTCTACTTCTGTATTAAGGCTGTGGTGGAGT 617
 OY 546 AGCATTAACAGATTCCTTCAGCGGATGATAAAGCTTCCGCTCGAGATCGACCA 605
 DB 618 AGCATTAACAGATTCCTTCAGCGGATGATAAAGCTTCCGCTCGAGATCGACCA 677
 OY 606 TGACAGAGTGAATCTCTGATTTAAATATGCTGTGACAGCATGATATATCTCTG 665
 DB 678 TGACAGAGTGAATCTCTGATTTAAATATGCTGTGATGATGAAATATATCTCTG 737
 OY 666 AAGGAGATTTTGGTATTAATCTTATGACATCATATGCTTTTCAATGTCAGATGC 725
 DB 738 AGGAGAGATTTTGGTATTAATCTTATGACATCATATGCTTTTCAATGTCAGATGC 797
 OY 726 TGACAGCAATTAATCTTGAAGCTCTCGACCATCATATCTCGTCTTCATCCAG 785
 DB 798 TGACAGCAATTAATCTTGAAGCTCTCGACCATCATATCTCGTCTTCATCCAG 857
 OY 786 AGAGAGATTTTCTTGTGGGCTGGAGAGACTTTAACTGTACAGATGATTTTAA 845
 DB 858 AGAAGAGATTTCTTGTGGGCTGGAGAGACTTTAACTGTATGATGATTTTAA 917
 OY 846 GTGAGAGAGAT 857
 DB 918 GTGGGAGAGAT 929

RESULT 7
 AV060640 684 bp mRNA linear EST 24-OCT-2001
 LOCUS AV060640 Mus musculus pancreas C57BL/6J adult Mus musculus CDNA
 DEFINITION AV060640.2 GI:16380782
 ACCESSION AV060640
 VERSION AV060640.2
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 684)
 AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koyama, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, T., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
 TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 JOURNAL Unpublished (2001)
 COMMENT On Jun 23, 1999 this sequence version replaced gi:5160387.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

100

1

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1058)				
NIH-MGC http://mgc.ncl.nih.gov/ .		National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	
Contact: Robert Strausberg, Ph.D.				
Email: cga@bts-remail.nih.gov				
Tissue Procurement: M. Bento Soares, Ph.D.				
cDNA Library Preparation: Bento Soares and M. Fatima Bonaldo				

QY 246 CATTTTGGGTCATTAAGTGTGTTGGGGTCAACATTTGAATAGATGCCACCAAG 305
 |||||||
 Db 275 CATTTTGGGTCATTAAGTGTGTTGGGGTCAACATTTGAATAGATGCCACCAAG 334
 |||||||
 QY 306 CTGCGACAGCAGCTGACACTTTCACAGCCAAAGTATGGATGCGGTCCAGAGATGAAT 365
 |||||||
 Db 335 CAGTACAGCAGCTGACACTTTCACAGCCAAAGTATGGATGCGGTCCAGAGATGAAT 394
 |||||||
 QY 366 TGTATGACCTTGGCTCATTAAGCAGATTTGCAAGACTGTGGATTTTCAACAGAGATGAAT 425
 |||||||
 Db 395 TGTATGACCTTGGCTCATTAAGCAGATTTGCAAGACTGTGGATTTTCAACAGAGATGAAT 454
 |||||||
 QY 426 ACTGNTAACTGGGGAGCAGATTAACCTGCTGCAATATGATTAACCAACCTGAG 485
 |||||||
 Db 455 ATTGTTTACCGGGGAGCAGATTAACCTGCTGCAATATGATTAACCAACCTGAG 514
 |||||||
 QY 486 CAGAACCTTAAGAAATAGTGGCCACACTTCTGTATTTAAAGAGCTGTGGTGCAGTG 545
 |||||||
 Db 515 CAGAACCTTAAGAAATAGTGGTCATCTTGTATTTAAAGAGCTGTGGTGCAGTG 574
 |||||||
 QY 546 AGCATTAACAGATCTTTCAGCGGATTAATTAACCTGTCGCTGCGATCAGCCAA 605
 |||||||
 Db 575 AGCATTAACAGATCTTTCAGCGGATTAATTAACCTGTCGCTGCGATCAGCCAA 634
 |||||||
 QY 606 TGACAGAGTGAATCTCTGATTTTAAATATGCTGTAGCAGATGAGATATTTCTG 665
 |||||||
 Db 635 TGACAGAGTGAATCTCTGATTTTAAATATGCTGTAGCAGATGAGATATTTCTG 694
 |||||||
 QY 666 AAGGAGAGATTTGGTATTAATTAAGTATGAGATCTTTCATAGTGCAGTAGTC 725
 |||||||
 Db 695 AAGGAGAGATTTGGTATTAATTAAGTATGAGATCTTTCATAGTGCAGTAGTC 754
 |||||||
 QY 726 TGAGCCAAATTAATCTTTCAGAGCTCTGCGACCATTAATGCTCTTTCATAGTCAG 785
 |||||||
 Db 755 TGAGCCAAATTAATCTTTCAGAGCTCTGCGACCATTAATGCTCTTTCATAGTCAG 814
 |||||||
 QY 786 AGAAGAGATTTGTTGTC--GGGTGAGAAAGCTTTAACTGATAGATGATTA 840
 |||||||
 Db 815 AGAAGAGATTTGTTGTC--GGGTGAGAAAGCTTTAACTGATAGATGATTA 871
 |||||||

RESULT 13
 B1687709
 LOCUS 926 bp mRNA linear EST 18-SBP-2001
 DEFINITION 6033156161 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:535866 5',
 mRNA sequence.
 ACCESSION B1687709
 VERSION B1687709.1 GI:15650337
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 926)
 TITLE NIH-MGC http://mgi.mc.man.ac.uk/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.fda.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/MLN at:
 http://image.llnl.gov
 Plate: L1M11904 row: o column: 03
 High quality sequence stop: 759.
 Location/Qualifiers
 1..926
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"

/clone="IMAGE:535866"
 /clone_1b="NCI CGAP Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DHI08"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site: 1; Salt:
 library constructed by Life Technologies. Primer: Oligo dT.
 providing samples: Jeffrey Green, M.D., NIH"
 BASE COUNT 276 a 190 c 199 g 260 t 1 others
 ORIGIN

Query Match 38.8%; Score 632.4; DB 13; Length 926;
 Best Local Similarity 98.3%; Pred. No. 1.1e-160;
 Matches 650; Conservative 0; Mismatches 7; Indels 4; Gaps 1;

QY 968 GTGTAGGAAAGACCTATGGCTGTGGAATGCGTGTCTGAGGAGACAGCGGGAA 1027
 |||||||
 Db 1 GTGTAGGAAAGACCTATGGCTGTGGAATGCGTGTCTGAGGAGACAGCGGGAA 60
 |||||||
 QY 1028 CTGGCAAGCCCAAGATCGGATTTCCAGAAACAGCAGAGAGAGCTGCGAGAGAAAT 1087
 |||||||
 Db 61 CTGGCAAGCCCAAGATCGGATTTCCAGAAACAGCAGAGAGAGCTGCGAGAGAAAT 116
 |||||||
 QY 1088 GCTTCAGAGATTCAGATTCATCTATTCATCAACCTCGAAGTTAAGCCCTGAGCATCA 1147
 |||||||
 Db 117 GCTTCAGAGATTCAGATTCATCTATTCATCAACCTCGAAGTTAAGCCCTGAGCATCA 176
 |||||||
 QY 1148 GACGTGCTGCCGAACCATATGTTGATGACATTAACAGCAGAGAGAGATCGGCT 1207
 |||||||
 Db 177 GACGTGCTGCCGAACCATATGTTGATGACATTAACAGCAGAGAGAGATCGGCT 236
 |||||||
 QY 1208 TCAGATTAAGTGTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
 |||||||
 Db 237 TCAGATTAAGTGTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 296
 |||||||
 QY 1268 TCAGATTAAGTGTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1327
 |||||||
 Db 297 TCAGATTAAGTGTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 356
 |||||||
 QY 1328 TGAGGAG 1387
 |||||||
 Db 357 TGAGGAG 416
 |||||||
 QY 1388 ACTTACATCTCCATTTTACACCTTAATTTCTTAAAGTATGATGAAGAGAA 1447
 |||||||
 Db 417 ACTTACATCTCCATTTTACACCTTAATTTCTTAAAGTATGATGAAGAGAA 476
 |||||||
 QY 1448 AATATATTGGCTATTTTGTGACTTCCCTTAAGAGAGAGAGAGAGAGAGAGAGAG 1507
 |||||||
 Db 477 AATATATTGGCTATTTTGTGACTTCCCTTAAGAGAGAGAGAGAGAGAGAGAGAG 536
 |||||||
 QY 1508 AGTATGAAG 1567
 |||||||
 Db 537 AGTATGAAG 596
 |||||||
 QY 1568 GGACTGCTTCAG 1627
 |||||||
 Db 597 GGACTGCTTCAG 656
 |||||||
 QY 1628 A 1628
 |||||||
 Db 657 A 657
 |||||||

RESULT 14
 B0439153
 LOCUS 875 bp mRNA linear EST 24-MAY-2002
 DEFINITION AGBNCOUFT_7827244 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6014262
 5', mRNA sequence.
 ACCESSION B0439153
 VERSION B0439153.1 GI:21178229

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 875)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM13208 row: h column: 07
High quality sequence stop: 675.
Location/Qualifiers
1. 875
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6014262"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 253 a 175 c 209 g 235 t 3 others
ORIGIN
Query Match 38.7%; Score 631.6; DB 14; Length 875;
Best Local Similarity 89.1%; Pred. No. 1.8e-160;
Matches 701; Conservative 0; Mismatches 84; Indels 2; Gaps 2;
QY 66 TCCTCCCTCCCTGTCGCCATGCGCATGAGGACACCCGCTGCTGCTGGGCCACA 125
DB 70 TCGCTGGCTTGCCTGCCCGCCATGCAATGACACACCCGCTGCTGCTGGGCCACA 129
QY 126 GCGGCGCCGCTGGTGTATGCGCTTACAGCGCATCAGCGCTTACGCTACTTCTGATCA 185
DB 130 CCGGACCCGCTGCTATTTGGCTTCAATGCGATCAGCGCTTATGGTATTTCTTAATCA 189
QY 186 GCGCTTGAAGAGTGAAGCCCATGCTCCGCCAGGAGATACAGAGACTGATTTGGA 245
DB 190 GCGCTTGAAGAGTGAAGCCCATGCTACGCGAGGAGATACAGAGACTGATTTGGA 249
QY 246 CATTTTGGGTATTAAGTGTGTTGGGGTGAACATTAAGTGAATGATGACCAAG 305
DB 250 CATTTTGGGTATTAAGTGTGTTGGGGTGAACATTAAGTGAATGATGACCAAG 309
QY 306 CTGCGACAGACCTCAGACTTCACAGCCAAAGTATGGAGTGGCTCTCAGAGATGAAT 365
DB 310 CAGCTACAGACCTCAGACTTCACAGCCAAAGTATGGAGTGGCTCTCAGAGATGAAT 369
QY 366 TGATGACCTGCTCATTAAGCAGATTTGTCAAGACTGTGATTTTCAACAGAGATGAAT 425
DB 370 TGATGACCTGCTCATTAAGCAGATTTGTCAAGACTGTGATTTTCAACAGAGATGAAT 429
QY 426 ACCTGTATCTGGGGAGAGATTAACCTGCTGCGCATATATGATGATGAACAACTGAAG 485
DB 430 ATTGTTTAAACCGGGAGACAGATTAACCTGTTACGCTATATGATGATGAACAACTGAAG 489
QY 486 CAGAACCTAAGAAATCACTGGCCACACTTGTGATTAAGAAAGCTCTGTGTGACAGTG 545
DB 490 CAGAACCTAAGAAATCACTGGCTATGCTGATTAAGAAAGCTCTGTGTGACAGTG 549
QY 546 ACATTAAGACATCTTTCAGCGGATGATTAAGAACTGTGGGCTCTGGGATTCATGCCACA 605

DB 550 AGATTAACAGATCTTTCGCTGATGACAAACTGTTCGCTTGGGATCATCTACTA 609
QY 606 TGACAGAGTAATCTCTGATTAATATGCTGTAGACAGATGAGATATTCCTG 665
DB 610 TGACAGAGTAATCTCTGATTAATATGCTGTAGAGATGAGATATTCCTG 669
QY 666 AAGGAGATTTTGTATTAATCTATGAGATGATTAATGCTTTTCAATAGTCAAGTGC 725
DB 670 AAGGAGATTTTGTATTAATCTATGAGATGATTAATGCTTTTCAATAGTCAAGTGC 729
QY 726 TGAGCCATTAATTC-TTTGAAGCTCTCGACCATTAATTCGCTCTNTTATCA 784
DB 730 TGAGCCATTAATTC-TTTGAAGCTCTCGACCATTAATTCGCTCTNTTATCA 789
QY 785 GAGAAAGATTTCTTGTGGGGTGAAGACATTT-AACTGTAAGATTAATTA 843
DB 790 GAGAAAGATTTCTTGTGGGGTGAAGACATTT-AACTGTAAGATTAATTA 849
QY 844 CAGTGA 850
DB 850 TAGTGA 856

RESULT 15
BI408095/LOCUS 1005 bp mRNA linear EST 14-AUG-2001
DEFINITION 602919077F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5055858 5',
mRNA sequence.
ACCESSION BI408095
VERSION BI408095.1 GI:15169018
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1005)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM1151 row: j column: 19
High quality sequence start: 15
High quality sequence stop: 845.
Location/Qualifiers
1. 1005
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5055858"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer (5',
TGTTACCAATCTGAAGTGGAGCGCGCTCTGTTTATTTTATTTTATTTT 3').
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia). digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT

289 a 229 c 185 g 302 t

ORIGIN

Query Match 38.7%; Score 631; DB 13; Length 1005;
Best Local Similarity 89.6%; Pred. No. 2.8e-160;
Matches 772; Conservative 0; Mismatches 71; Indels 19; Gaps 8;

QY 781 TCCAGAGAAAGAGTTCTGTTGGGGGTGGAGAACTTTAACTGTACAGTATGATTA 840
DB 1004 TTCAATCCAGAGAAAGATTTCTTTTGGGTGGAAAACTTTAACCTGGACAGTATGATTA 945
QY 841 TAACAGTGGAGAA--GAGTTAGAACTCTACAAAGGTCACTTTGGTCCATTCACTGT--GT 897
DB 944 TAGCAGTGGAGAAAGAGTTGAATCCCTACAAAGTCCCTTTGGTCCATTCACTGTGG 885
QY 898 GAGATTGAGTCTGATGGGGAACCTAT--GCCAGCGGTTCTGAAGATGGGACATTGAG 954
DB 884 GAGATTGAGTCTGATGGGGAACCTAT--GCCAGCGGTTCTGAAGATGGGACATTGAG 825
QY 955 A-----TTGTGGCAAACTGTGTAGAAAGAACTATGGCTGTGGAA--TGCGTGT 1006
DB 824 AGATTGTGGCAAACTGTGTAGAAAGAACTATGGCTGTGGAAATGCGTGTTC 765
QY 1007 CCTGAGAGAGACAGCGGGGAACCTGGCAAAAGATCGGATTTCCAGAAACAGCAGAG 1066
DB 764 CCTGAGAGAGACAGCGGGGAACCTGGCAAAAGC--AAGATCGGATTTCCAGAAACAGCAGAG 706
QY 1067 GAAGAGTGGCAGAAAGAAATTTGCTCAGAGAAATTCAGATTCCATCTATTCACACTCT 1126
DB 705 GAAGAGTGGCAGAAAGAAATTTGCTCAGAGAAATTCAGATTCCATCTATTCACACTCT 648
QY 1127 GAAGTTAAGGCTGAGCATCAGAGTGTGCTGCCAAACCATATGTCATGAGACTAACA 1186
DB 647 GAAGTTAAGGCTGAGCATCAGAGTGTGCTGCCAAACCATATGTCATGAGACTAACA 588
QY 1187 AGCAGAGAAAGCATCCGCTCTCAGAGTACTGTCTGCTGAGGCAAAAGAGGCGAGAAA 1246
DB 587 AGCAGAGAAAGCATCCGCTCTCAGAGTACTGTCTGCTGAGGCAAAAGAGGCGAGAAA 528
QY 1247 TATTGGGCAATATGATTTAGTTCAGTGCAGAAACAGTACTAGTGTGCGCGTGAAGT 1306
DB 527 TATTGGGCAATATGATTTAGTTCAGTGCAGAAACAGTACTAGTGTGCGCGTGAAGT 468
QY 1307 AAAATGGCTAGTGTCTGAGGTGAGGAGAGAGATTTGCTCAGATAGTGCATAGCT 1366
DB 467 AAAATGGCTAGTGTCTGAGGTGAGGAGAGATTTGCTCAGATAGTGCATAGCT 408
QY 1367 GCTGTTGGAATGAAAAGCAACTTACATCTCCATTTTACACCTAAATTTCTTTAGCT 1426
DB 407 GCTGTTGGAATGAAAAGCA--CTTACATCTCCATTTTACACCTAAATTTCTTTAGCT 350
QY 1427 GTTTATGTTATGAGAGAAATATATTGGCTATTTTCTGACTTCCCTTAAGAAG 1486
DB 349 GTTTATGTTATGAGAGAAATATATTGGCTATTTTCTGACTTCCCTTAAGAAG 290
QY 1487 AATGCTTTTGTCTGCTGCTAGTATGAAGAGAGAAATACATGATAAAGTAAACGGT 1546
DB 289 AATGCTTTTGTCTGCTGCTAGTATGAAGAGAGAAATACATGATAAAGTAAACGGT 230
QY 1547 TTGATCTCTTTCATTTACAAAGAGCTGCTCAGAAAGCTCAATTTTATGTTATCTAAA 1606
DB 229 TTGATCTCTTTCATTTACAAAGAGCTGCTCAGAAAGCTCAATTTTATGTTATCTAAA 170
QY 1607 TAAATATGCTTAAATATTAATA 1628
DB 169 TAAATATGCTTAAATATTAATA 148

Search completed: February 6, 2003, 13:03:48
Job time : 2471 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 6, 2003, 13:05:33 ; Search time 3249 Seconds

(without alignments)
3144.070 Million cell updates/sec

Title: US-09-856-836-2
Sequence: 1 MAMQTPITCSGHRPVVDX.....EETASSENSDITSTPEVKA 351

Scoring table:

BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q/cg2_1/USFTO.spool/US0856836/runat_31012003_144237_22195/app_query.fasta_1.519
-DB=GenBml -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPT=0 -LOOPTXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsm62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=2000000000
-USRR=US0856836.ecgn 1.1616.0runat_31012003_144237_22195 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -IARECOVERY -NRC_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

GenBml: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pla: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vtl: *
38: em_sy: *
39: em_htg_hum: *
40: em_htg_mus: *
41: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1832	99.6	1630	10 AF096285	AF096285 Mus muscu
2	1784.5	97.0	1563	9 AY049776	AY049776 Homo sapi
3	1784.5	97.0	1712	9 BC000162	BC000162 Homo sapi
4	1784.5	97.0	1716	9 HAUI0025	AJ010025 Homo sapi
5	1780.5	96.8	1867	9 HSM801659	ALJ36691 Homo sapi
6	1779.5	96.7	1820	6 AR053395	AR053395 Sequence
7	1777.5	96.6	1864	9 AB024327	AB024327 Homo sapi
8	1548.5	84.2	154115	9 AC010878	AB024327 Homo sapi
9	1418	77.1	1829	5 AF315726	AF315726 Homo sapi
10	1334.5	72.5	187017	2 AC010101	AF315726 Carassius
11	1328.5	72.2	166680	2 AC010132	AC010101 Homo sapi
12	1128	61.3	73038	2 AC020884_3	AC010132 Homo sapi
13	1128	61.3	226520	10 AL645802	Continuation (4 of
14	1062	57.7	170393	10 AY061000	AL645802 Mouse DNA
15	1052.5	57.2	1591	3 AY061000	AC109605 Mus muscu
16	876.5	47.6	31772	2 AC020469	AC020469 Drosophila
17	876.5	47.6	188272	3 AC005639	AC005639 Drosophila
18	876.5	47.6	218565	3 AC009257	AC009257 Drosophila
19	876.5	47.6	295225	3 AE003461	AE003461 Drosophila
20	781.5	42.5	1411	8 AY084604	AY084604 Arabidops
21	766.5	41.7	1399	8 AY074392	AY074392 Arabidops
22	549.5	29.9	81370	8 AB017071	AB017071 Arabidops
23	533	29.0	413	9 HS056430	US6430 Human HeLa
24	517.5	28.1	313	6 AX261089	AX261089 Sequence
25	506	27.5	91001	8 AC013453	AC013453 Arabidops
26	506	27.5	141984	8 F9L1	AC007591 Arabidops
27	489	26.6	98151	2 AC119404	AC119404 Magnapor
28	487	26.5	311	6 AC0393751	AX393751 Sequence
29	477	25.9	34410	3 AC008242	AC008242 Leishmani
30	477	25.9	116000	2 AC129714	AC129714 Leishmani
31	454.5	24.7	89765	2 AC017870	AC017870 Drosophila
32	454.5	24.7	188873	3 AC009009	AC009009 Drosophila
33	454.5	24.7	196337	3 AC005894	AC005894 Drosophila
34	454.5	24.7	242172	3 AE003832	AE003832 Drosophila
35	452	24.6	117737	8 F6D8	AC008016 Arabidops
36	452	24.6	126253	8 AC019018	AC019018 Arabidops
37	425	23.1	686	6 AR144171	AR144171 Sequence
38	425	23.1	686	6 AR176358	AR176358 Sequence
39	425	23.1	686	6 AR176358	AR176358 Sequence
40	394	21.4	10048	3 AY061931	AY061931 Drosophila
41	366.5	19.9	1200	8 AF285835	AF285835 Arabidops
42	366.5	19.9	1205	8 AF370485	AF370485 Arabidops
43	363.5	19.8	1011	8 AY064633	AY064633 Arabidops
44	347.5	18.9	1138	8 ATU36765	U36765 Arabidops
45	344.5	18.7	1143	8 AF335551	AF335551 PhaseoLus

RESULT 1

ALIGNMENTS

AF096285 1630 bp mRNA linear ROD 27-DEC-1998
 LOCUS AF096285
 DEFINITION Mus musculus serine-threonine kinase receptor-associated protein
 mRNA, complete cds.
 ACCESSION AF096285
 VERSION AF096285.1 GI:4063382
 KEYWORDS
 ORGANISM Mus musculus.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1630)
 AUTHORS Datta, P. K., Chytil, A., Gorska, A. E. and Moses, H. L.
 TITLE Identification of STRAP, a novel WD domain protein in transforming
 growth factor-beta signaling
 J. Biol. Chem. 273 (52), 34671-34674 (1998)
 JOURNAL 99074230
 MEDLINE 9856985
 PUBMED 9856985
 REFERENCE 2 (bases 1 to 1630)
 AUTHORS Datta, P. K.
 TITLE Direct Submission
 Submitted (01-OCT-1998) The Vanderbilt Cancer Center, Vanderbilt
 University School of Medicine, 649 Medical Research Building II,
 Nashville, TN 37232-6838, USA
 JOURNAL location/Qualifiers
 FEATURES
 source 1. 1630
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /dev_stage="embryo"
 86. 1141
 /note="WD domain protein; STRAP; TGF-beta
 receptor-associated protein"
 /codon_start=1
 /product="serine-threonine kinase receptor-associated
 protein"
 /protein_id="AAC98300.1"
 /db_xref="GI:4063383"
 /translation="MAMKQTEPLTCSGHTRPVVDLAFSGITPYGYFLISACKDQKPMRL
 OGDWDVIGFGLGKAWGATLNDKATKATADAFKAWDAVSGDELMTLAKHAI
 VKTVDFPDSDNYLTGGODKLRTYDLKPEAPREIGHTSGIKKALMSDDQILS
 ADKTVPLDMHATVETKSLINFMVSSEMEYIPGEILVITYGSIHAFSAVLEPIK
 SFEPATINASLSPEKEFLVAGGEDFKLYKDYNSGELEFSYGHREGHICVRSPP
 GELVASSEDGTLKLMQTVVGTIKYGLMKCVLPEDSGELAPKIGFPETADELAELI
 ASENSDSYSSSTPEVKA"
 BASE COUNT 455 a 349 c 406 g 420 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,69e-158 Length: 1630
 Score: 1832.00 Matches: 347
 Percent Similarity: 98.86% Conservative: 0
 Best Local Similarity: 98.86% Mismatches: 4
 Query Match: 99.57% Indels: 0
 DB: 10 Gaps: 0
 US-09-856-836-2 (1-351) x AF096285 (1-1630)
 QY 1 MetAlaMetArgInThrProLeuThrCysSerGlyHisThrArgProValAsp*** 20
 DB 86 ATGGCATTAGGACAGCGCGCTCATCTGCGGCACACGGCGCGGTGGATTG 145
 QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 146 GCCTTCAGCGGACATCAGCGCTTACGCTTCTCATCAGCGCTTGCAAAAGATGGCAAG 205
 QY 41 PromELeuArgGlnGlnAspThrGlyAspThrIleGlyThrPheLeuGlyHisLysGly 60
 DB 206 CCATCTCTCCGCGGAGGATACAGGAGCTGATTCGAAATTTTGGTCTATAAGGT 265
 QY 61 AlaValTrrGlyAlaThrLeuAsnLysAspAlaThrLysAlaIleThrAlaAlaIleAsp 80
 DB 266 GCTGTTGGGTCACATTTGATTAAGATGCCAACCAAGCTCGACAGCAGCTGCAGAC 325

QY 81 PheThrAlaLysValTrrPAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 DB 326 TTCACAGCCAAAGATGTGGATGCGCTCTCAGACAGATGATTAAGATGACCCCTGCTCATTAAG 385
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu***ThrGlyGlyGln 120
 DB 386 CACATTCGCAAGACTGTGATTTCCACACAGATAGCAATTCCTGCTAACTGGGGGACAG 445
 QY 121 AspLysLeuArgGlyIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
 DB 446 GATTAACCTGCGCATATATGACTTTCAGACAACTGAGACAGAACTCAAGAAATACGT 505
 QY 141 GlyHisThrSerGlyIleLysAlaLeuThrCysSerAspAspLysGlnIleLeuSer 160
 DB 506 GGCCACACTCTGGATTAATAAAGCGCTGTGGTCAGATGACATTAATTCGTAAGATTTGGTTAT 565
 QY 161 AlaAspAspLysThrValArgLeuThrAspHisAlaThrMetThrGluValLysSerLeu 180
 DB 566 GCGGATGATTAACCTGTCGCTGCGATCATGCGCAATGACAGAAAGATTAATCTCTG 625
 QY 181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyGluIleLeuValIle 200
 DB 626 AATTTATATCTCTGTTAGCAGCATGAGATATATTCGTAAGAGAGATTTGGTTAT 685
 QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSerPhe 220
 DB 686 ACTTATGACAGCATATATGCTTTTCATAGTGCATGATGATGAGCAATTAATTCCTTT 745
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGluLysGluIleLeuValAla 240
 DB 746 GAACTCTCTGCGACCATCATATTCGCTCTTCATCCAGAAAGAGATTTTGTGG 805
 QY 241 GlyIleGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluIleLeuGluSer 260
 DB 806 GCGTGAAGAACTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 865
 QY 261 TyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspLysLysLeuTyr 280
 DB 866 TACAAGGCTCTGTTGCTCCATTCATCTGTGACATTAAGTCTGATGAGGAGACTCTAT 925
 QY 281 AlaSerGlySerGluAspLysThrLeuArgLeuTrrPglThrValAlaGlyLysThrTyr 300
 DB 926 GCGAGCGTTCGAAAGATGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 985
 QY 301 GlyLeuThrLysCysVal***ProGluLysAspSerGlyGluLeuAlaLysProLysIle 320
 DB 986 GGCCTGTGGAATGCGCTGCTCTGAGGAAGACAGCGGGAACCTGCAAGCCAAAGATC 1045
 QY 321 GlyPheProGluThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluAsnSerAsp 340
 DB 1046 GGAATTCGAAAGAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1105
 QY 341 SerIleTyrSerSerThrProGluValLysAla 351
 DB 1106 TCCATCTATTCATCACTCTGTAAGTAAAGGCC 1138
 RESULT 2
 AY049776 1563 bp mRNA linear PRI 23-OCT-2001
 LOCUS AY049776
 DEFINITION Homo sapiens unr-interacting protein (UNRIP) mRNA, complete cds.
 ACCESSION AY049776
 VERSION AY049776.1 GI:16356636
 KEYWORDS
 ORGANISM Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1563)
 AUTHORS Liu, J., Zhou, Y., Peng, X., Yuan, J. and Qiang, B.
 TITLE Direct Submission
 Submitted (31-JUL-2001) Department of Biochemistry, Institute of
 Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong
 Dan San Tiao, Beijing 100005, P.R. China


```

FEATURES
  source      1..1563
              Location/Qualifiers
                1..1563
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                1..1563
                /gene="UNRIP"
                48..1100
                /gene="UNRIP"
                /codon_start=1
                /product="unr-interacting protein"
                /protein_id="AA15433.1"
                /db_xref="GI:1635637"
                /translation="MAMQRPPLTSGHTRPVVDLAFSGITPYGFLISACKDGRPMRL
                OGCDMDGATGFLGKAWGATLTKNDATKATTAADLPAKWDVSGDELTAAHAKHI
                VKIVDFDSDNATLTLGGQDKLRIYDINKPEAPEKEISGHTSGIKKALWCSEDKOIS
                ADEPTRLMDATMTAEVKSILNFMNSVSMEXIPEGEILIVTIGRSIAFHSVSLDPIK
                SFEPARTNSALPEKEFLVAGGEDFLYADYDNGSELESTYKGHFCHVCRSPD
                GELVSGSEDTLRIMQTVGVKTYGLMKCVLPEDSBLAKPKIGFTTTELELETA
                SENSDCIFPSAPDYKA"
BASE COUNT   465 a      284 c      362 g      452 t
ORIGIN
Alignment Scores:
  Pred. No.:      1,38e-153      Length:      1563
  Score:          1784.50         Matches:      338
  Percent Similarity: 97.44%      Conservative: 4
  Best Local Similarity: 96.30%   Mismatches: 8
  Query Match:     96.98%         Indels:      1
  DB:               Gaps:         1
US-09-856-836-2 (1-351) x AY049776 (1-1563)

OY 1 MetAlaMetAArgGlnThrProLeuThrcysSerGlyHisThrArgProValAlaAsp*** 20
Db 48 ATGGCAATGAACAGACGCCCTCACTGCTGCGCCACAGCAGCCGCTGTCATTG 107
OY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
Db 108 GCCTTCAGTGCATCAGCCCTTATGGGATTTCTTAATCAGCGCTTGCAAAAGATGTAA 167
OY 41 ProMetLeuArgGlnGlyAspThrGlyAspTrrIleGlyThrPheLeuGlyHisLysGly 60
Db 168 CCAATGTCACGCCAGGAGATACAGGACACGATGGAACATTTTGGCTCAATAAGGT 227
OY 61 AlaValTrrGlyAlaThrIleuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
Db 228 GCTGTTGGGGTGCACACTGAATGAAGATGCCACCAAGCAGCTACAGCAGTCAGAT 287
OY 81 PheThrAlaLysValTrrPaspAlaValSerGlyAspGluLeuMetThrIleuAlaHisLys 100
Db 288 TTCACAGCCAAAGTGGGATGCTGCTCAGAGAGATGAATGATGACCCGCTCATMAA 347
OY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrIleu***ThrGlyGlyGln 120
Db 348 CACATTTTCAGAGCTGTGGATTTTCACGAGATGAATATTAATTTGTTAAACCGGGGACAG 407
OY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
Db 408 GATAAATCTGTACGATATATGACTTGAACAAACCTGAACAGACCTTAAGCAATTAAGT 467
OY 141 GlyHisThrSerGlyIleLysLysAlaLeuTrrCysSerAspLysGlnIleLeuSer 160
Db 468 GGTCAATACCTCTGATTAATAAAGCTGTGTGTCAGATGAGATTAACAGATTTCTTCT 527
OY 161 AlaAspAspLysThrValArgLeuTrrPaspHisAlaThrMetThrGluValLysSerLeu 180
Db 528 GGTGATGACAAACCTGTGCACCTTGGGATCAGCTACTACTACAGAGGGAATGTCTA 587
OY 181 AsnPheAsnMetSerValSerMetGluTyrIleProGluGlyLysIleLeuValIle 200
Db 588 AATTTTAATATGCTCTGTAGTATGAAATATATCTCGAGGAGAGATTTTGGTTATA 647
OY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220

648 ACTTATGACGACATCATTCCTTTTCATATGTCAGTAACTTGGACCCATTAATACCTTT 707
OY 221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
Db 708 GAAGCTCCTCCACACCATCAATTCGATCTCTCATCTCGTGAAGAAGATTTCTTGTTCA 767
OY 241 GlyGlyLysAspPheLysLeuTyrLysTyrAspPyrLysAsnSerGlyGluGluLeuSer 260
Db 768 GCGGTGAGAAATTTTAAACCTTATAGTATGATATATATATGAGAGAAATTTAGAAATTC 827
OY 261 TyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyLysLeuTyr 280
Db 828 TACAAGGACACCTTTGGTCCATTCACCTGCTGATTTAGTCTGATGAGAACTCTAT 887
OY 281 AlaSerGlySerGluAspGlyThrLeuArgLeuTrrPdpInThrValAlaGlyLysThrTyr 300
Db 888 GCCAGTGTTCAGAAAGATGGAACATGAGACTATGCAACACTGCGTAGGAAAAACGTAT 947
OY 301 GlyLeuTrrPysCysVal***ProGluLysAspSerGlyLysLeuAlaLysProLysIle 320
Db 948 GGCCTTTGGAATATGTGCTTCTCGAAGAAAGATAGTGTAAGCTGGCAAGCCAAAGATT 1007
OY 321 GlyPheProGluThrAlaGluGluGluLysLeuAlaGluLysIleAlaSerGluAsnSerAsp 340
Db 1008 GGTTCACAGACACAAACAGAAAGAGAGCTA---GAAGAATTCCTTCAGAGAAATTCAGAT 1064
OY 341 SerIleTyrSerSerThrProGluValLysAla 351
Db 1065 TGCATCTTCTTCACGCTCGATGTAAAGGCC 1097

RESULT 3
BC000162      1712 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS
DEFINITION
Homo sapiens, unr-interacting protein, clone MGC:5236
IMAGE:2800954, mRNA, complete cds.
ACCESSION
BC000162
VERSION
BC000162.2 GI:14198026
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1712)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (03-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
On May 25, 2001 this sequence version replaced gi:12652818.
COMMENT
Contact: MGC help desk
Email: gcgphs.rem@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: Villalobos@bcm.tmc.edu.
Villalobos, D.R., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IMAGE Plate: 3 Row: e Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6807652.
Location/Qualifiers
1..1712
FEATURES
  source

```

```

/organism="Homo sapiens"
/db_xref="locusid:11171"
/db_xref="taxon:9606"
/clone="MGC:5236 IMAGE:2900954"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NH MGC_10"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
146. .1198 -1

```

BASE COUNT	516 a	319 c	395 g	482 t
ORIGIN				

Alignment Scores:	1,54e-153	1,712
Pred. No.:	1,784	338
Score:	97,44.50	
Percent Similarity:	96.308	
Best Local Similarity:	96.988	
Query Match:	9	1
DB:	9	1
	Gaps:	
	Indels:	
	Matches:	
	Conservative:	
	Mismatches:	
	Length:	

US-09-856-836-2 (1-351) x BC000162 (1-1712)

OY	1	MetAlaMetAlaGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp***	20
Db	146	ATGGCAATGAGACACAGACGGCGCTCACCCTGTGGGCACACGGGACCCGTGGTTGATTTGG	205
OY	21	AlaPheSerGlyIleIleThrProTyrGlyTyrPheIleuIleSerLacCysLysAspGlyLys	40
Db	206	GCCCTTCAGTGGCATTACAGCCCTTATGGGTATTCTTATACAGCCCTTGCAAAAGATGTAA	265
OY	41	PrometIleuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheIleuGlyHisLysGly	60
Db	266	CCTATGCTACCCCGACGGGAGATACAGGAGACTGGATGGCAATTTTGGCTCTAAAGGT	325
OY	61	AlaValIleTyrGlyAlaThrIleuAsnLysAspAlaThrLysAlaIleThrAlaIleAlaAsp	80
Db	326	GCTGTTTGGGTGCCACACCTAATTAAGATGCCACCAACAGACTACAGCAGCTGCAGAT	385
OY	81	PheThrAlaLysValTyrPaspAlaValSerGlyAspCJuleuMetThrIleuAlaHisLys	100
Db	386	TTCCACAGCCAAAGTGTGGATGTGTCTCAGAGAAATGAAATGTATGACCCCTGCTCATAA	445
OY	101	HisIleValLysThrValAspPheThrClnAspSerAsnTyrIleu***ThrGlyLysGln	120
Db	446	CACATTGTCAAGACACTGGATTTTCACGAGAGATGTAATTAATTTGTTAAACGGGGGACAG	505
OY	121	AspLysIleuLeuArgIleTyrAspLeuAsnLysProGluIleGluProLysGluIleSer	140
Db	506	GATTAACCTGTTACGATATTATGACTTGACAAACACTGACAGCAGAACTAAGGAAATTACT	565
OY	141	GlyHisThrSerGlyIleLysLysAlaLeuThrPcysSerAspAspLysGlnIleLeuSer	160
Db	566	GGTCATTACTTGGTATTAATAAAACCTCTGTGGTGAGTGAGTGAATAACGATTCCTTCT	625
OY	161	AlaAspAspLysThrValArgLeuThrPaspHisAlaThrMetThrGlyIleuLysSerLeu	180
Db	626	GGCGATGACAAACACTGGTTCGACTTTGGGATCATGCTACTAAGACAGAGGAAATTCCTA	685
OY	181	AsnPheAsnMetSerValSerSerMetLysIleProGluGlyGluIleLeuValIle	200
Db	686	AATTTTAATATCTCTGTTACTGATGCAATATATCTCTGAGGAGAGATTTTGGTTATA	745

Oy	201	ThryrgrglAagSerlleAlaPhHisSerLalValSerLeuglupProilleYserPhe	220
Dd	746	ACTTATGGACGACATCATTCGTCTTTCAATAGTCACAGTAATTGGACCATAATCCTTT	805
Oy	221	GluAlaPrAlarHrllleAsnSerAlaser**HisProgluLySgluPheLeuValAla	240
Dd	806	GAAGCTCCTGCACACCATCAATTCGACATCTTCATCCCTCGAGAAGAATTCCTGTGGCA	865
Oy	241	GlyglYgluAsppHeuLeuTyrllysTrsprHyrsnsrgLYgluLdeuGlser	260
Dd	866	GGCGGTGGAAGATTTTAAACTTATDAAGTATATTAATTAAGGAGAGAAATTGGAATCC	925
Oy	261	TyrLySGlYhASpHeGlProiliehiscysValArqPheSerProaspGlygluLeuTyr	280
Dd	926	TACAAGGGACACTTGTCCTATTCACGTGTGNAGATTTAGCCGATGAGAGAACTCAT	985
Oy	281	AlaseRgLySerGluAspGlyThrIeuargleutrpGlnThrValGlyLyThrTyr	300
Dd	986	GCCACTGTGTTCCGAAGAATGGAACATTTGACATATGGCAAACGTGCGVGAAGAAAAAGTAT	1045
Oy	301	GlyLeuTrpLyScysVal**ProgluLjuaSpSerGlygluLeuAlaLySProlsile	320
Dd	1046	GGCCCTTTGGAAATGTGTCTTCCTTAAGAACATATGTGGTGCAGCTGGCAAGCCAAGATT	1105
Oy	321	GlyPheTrpGluThrAlaglucLuLeueualaglujunlIleAlaSerGluasnSerasp	340
Dd	1106	GGTTTTCCAGACACAACAGNAGAGAGCTA--GAAGAATATGCTTCAGAGATATCACAT	1162
Oy	341	SerIleTyrSerSerThrProgluVallyslala	351
Dd	1163	TGCATCTTTCCTGCAGCTCTGATGTAAAGGCC	1195

RESULT	4
HAIJ0025	
LOCUS	HAJJ0025 1716 bp mRNA linear PRI 12-MAR-1999
DEFINITION	Homo sapiens mRNA for unr-interacting protein.
ACCESSION	AJ010025
VERSION	AJ010025.1 GI:4468217
KEYWORDS	RNA-binding protein; unr-interacting protein; unr1p gene.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 1716)
AUTHORS	Hunt,S.L.
TITLE	Direct Submission
JOURNAL	Submitted (06-AUG-1998) Hunt S.L., University of Cambridge, Department of Biochemistry, 80 Tennis Court Road, Old Addenbrooke's Site, Cambridge, CB2 1GA, U.K
REFERENCE	2 (bases 1 to 1716)
AUTHORS	Hunt,S.L., Hsuau,J.J., Totty,N. and Jackson,R.J.
TITLE	unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA
JOURNAL	Genes Dev. 13 (4), 437-448 (1999)
MEDLINE	99158740
PUBMED	10049359
FEATURES	
source	Location/Qualifiers
	1..1716

FEATURES	
source	location/Qualifiers
gene	1..1716
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_line="Hela cell"
	148..1200
	/gene="unrlp"
	148..1200
	/gene="unrlp"
	/function="Interacts with human cold shock domain protein, unrlp"
	/codon_start=1
	/evidence=experimental
	/product="unrl-interacting protein"
	/protein_id="CAB38041.1"
	/db_xref="GI:4468218"

/translation="MAMROPELTCGSHRPVYDLAFSGITPYGELISACKDGRMLR
 OGDGWDITGFLGKAGWAGTFLNKDAFKATTAADFLAKWDAVSGDEMTLAKHL
 VKTYDFPDSDATLTLGGODKLLRTYDLKPEAREKISGHTSGIKRLMCSKQIIS
 ADKTYRLMDHATMEKSLNFSNMSVSMEYIPGELIVYIGSIFHSAVSDPIK
 SEAPATINSAHLEPEKEFLVAGEDFLYDYDNGSELESYSGHGPICVPSPD
 GELVASEGSDTLRLQTVVGTGYGLMWCVLPEDSDGLAPKICPETTBELEELA
 SENSDFIPSPADVKA"

ORIGIN 519 a 320 c 395 g 482 t

Alignment Scores:

Pred. No.:	1,54e-153	Length:	1716
Score:	1784.50	Matches:	338
Percent Similarity:	97.44%	Conservative:	4
Best Local Similarity:	96.30%	Mismatches:	8
Query Match:	96.98%	Indels:	1
DB:	9	Gaps:	1

US-09-856-836-2 (1-351) x HAM10025 (1-1716)

QY 1 MetalameAargInThProleuThrcysSerGlyHisThrArgProValAsp*** 20
 DB 148 ATGGCAATGACAGACAGACCCCTCCTCCTGCTGACAGCAGCAGCCCGTGGTGAATTG 207
 QY 21 AlaPheSerGlyIleThrProtyrGlyTyrrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 208 GCCTTCAGTGCATCAGCCCTTATGGGATTTCTTAATCAGCCGCTGCAAGATGGTAA 267
 QY 41 ProMetLeuAargInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 DB 268 CCTATGCTACCCAGGAGATACAGAGACAGATGGATTTGGTGGCTCATTAAGGT 327
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaIleThrAlaAlaAsp 80
 DB 328 GCTGTTGGGGTGCACACTGAATTAAGATGCCACCAAGCAGCTACAGCAGCTGACAT 387
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGlyLeuMetThrLeuAlaHisLys 100
 DB 388 TTCACACCAAGAGTGGGATGCTGCTCAGAGATGAATTTGATGACCTGGCTCATTA 447
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrrLeu***ThrGlyGln 120
 DB 448 CACATTTCAAGACTGTCGATTTCCACCAAGATGATTAATTTGTTAAACCGGGGACAG 507
 QY 121 AspLysLeuLeuArgIleTyrrAspLeuAsnLysProGlnIleGluProLysGlyLysSer 140
 DB 508 GATTAACCTGTACGATATGACTTGAACAAACCTGACAGCAGACCTAAGCAATTAGT 567
 QY 141 GlnHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSer 160
 DB 568 GGTCTACTTCTGGTATTAATAAAGCTCTGTGTCAGTGAAGATTAACAGATTTCTTCT 627
 QY 161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeu 180
 DB 628 GCGATGACAAACCTGTCGACTTGGATCAGCTACTATGACAGAGTGAAGATCTCTA 687
 QY 181 AspPheAsnMetSerValSerSerMetGluTyrrIleProGlnIleGluLeuValIle 200
 DB 688 AATTTTAATATGCTGTAGTATGATGAATATATCTGAGGAGAGATTTGTTATTA 747
 QY 201 ThrTyrrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
 DB 748 ACTTATGAGCAGATCTATTTGCTTTTCAATAGGTCAGTGAAGTTCGACCAATTAATCTT 807
 QY 221 GlnAlaProAlaThrIleAsnSerAlaSer***HisProGlnIleGluPheLeuValAla 240
 DB 808 GAAGCTCTGCAACATCAATCTGCATCTTCATCCTGCAAGAAACAAATTTCTGTCGA 867
 QY 241 GlnGlyGlnAspPheLysLeuTyrrLysTyrrAspTyrrAsnSerGlyGluLeuGluSer 260
 DB 868 GCGGCGTGAAGATTTTAACCTTATTAAGTATGATTAATATGAGAGAAAGATTAAGATCC 927
 QY 261 TyrrLysGlnHisPheGlyProIleHisCysValArgPheSerProAspGlyLeuLeuTyrr 280

DB 928 TACAAGGACACTTGGCTATTACAGTGTGAGATTGATGAGAACTTAT 987
 QY 281 AlaSerGlySerGluAspGlyThrLeuArgLeuTrpIleThrValAlaGlyLysThrTyrr 300
 DB 988 GCGAGTGTTCAGAAAGATGCAACATTGAGACTAATGCGCAACCTGCTGAGCAAAACCTAT 1047
 QY 301 GlyLeuTrpLysCysVal***ProGlnIleAspSerGlyGluLeuAlaLysProLysIle 320
 DB 1048 GGCCTTGGAAATGTGTGCTTCCGAAAGAGATGTGTGAGCTGGCAAGCCAAAGATT 1107
 QY 321 GlyPheProGlnThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluAsnSerAsp 340
 DB 1108 GGTTCCTCAGAGACACAGCAAGAGAGAGCTA---GAAGCAATTCCTTCAGAGATTCAGAT 1164
 QY 341 SerIleTyrrSerSerThrProGluValLysAla 351
 DB 1165 TGCACTTTCCTTCACCTCCTGATGTTAAGGCC 1197

RESULT 5
 HSM801659
 LOCUS
 DEFINITION
 complete cds.
 accession
 version
 keywords
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1867)
 Wiemann,S., Well,B., Wellenreuther,R., Gassenhuber,J., Glasl,S.,
 Ansoerge,W., Boecker,M., Bloeker,H., Baerach,S., Blum,H.,
 Lauber,J., Duesterhoeft,A., Beyer,A., Koehrer,K., Strack,N.,
 Mewes,H.W., Ottenwaelder,B., Obermaier,B., Tampe,U., Heubner,D.,
 Toward a Catalog of Human Genes and Proteins: Sequencing and
 Analysis of 500 Novel Complete Protein Coding Human cDNAs
 Genome Res. 11 (3), 422-435 (2001)
 11230166
 2 (bases 1 to 1867)
 Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
 Wiemann,S.
 Direct Submision
 Submitted (15-JAN-2000) MIPS, Am Klopfersplitz 18a, D-82152
 Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by MedGenomix (Martinsried/Germany) within the cDNA
 sequencing consortium of the German Genome Project. This clone
 (DKFZps64N1778) is available at the RZPD in Berlin. Please contact
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
 source
 location/Qualifiers
 1..1867
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZps64N1778"
 /tissue_type="brain"
 /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host
 X1-2blue; sites NotI + SalI"
 /dev_stage="fetal"
 297..1349
 /gene="DKFZps64N1778"
 297..1349
 /gene="DKFZps64N1778"
 /note="unr-interacting protein"
 /product="hypoetical protein"
 /protein_id="CAB6626.1"

CDs

/db_xref="GI:6807653"
 /db_xref="SPRMBL:09NTK0"
 /translation="MAMROTPLTSCGHTTRPVVDLARSITPYGYFLISACKDKPMIR
 OGDTWIGTFLHKGAWGATLNDKATKAATAAFATKAWMVAVSDELMIAHKHI
 VKTVDPTDSNYLITGGOKLRIDYNKEAPETISGIRKALMCSEROLIS
 ADDKYATLMDHATMEYKSLNPMYSVSMETEGELIVTVGRSLAFHSAVSLDPK
 SEDPATINSAHPEKLEVLAGEDEPKLYKTYINSESELESTKNGHGRTHCREFSD
 GELVAGSEDEGLRMOTVYGVKTYGLMKCVLPEDSEGLAKPKIPETTEEBLEELA
 SENSEIFFSAPVKA"
 1829. .1834
 polyA_signal
 1842
 polyA_site 532 a 388 c 422 g 525 t
 BASE COUNT
 ORIGIN

Alignment Scores:
 Pred. No.: 3,95e-153 Length: 1867
 Score: 1780.50 Matches: 337
 Percent Similarity: 97.44% Conservative: 5
 Best Local Similarity: 96.01% Mismatches: 8
 Query Match: 96.77% Indels: 1
 DB: Gaps: 1

US-09-856-836-2 (1-351) x HSM801659 (1-1867)

QY 1 MetAlaMetArgInThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
 Db 297 ATGGCAATGAGACAGACCGCCGCTACCTGCTCTGGCCACGCGACCCGCTGATTG 356
 QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 Db 357 GCCTTCATGCGATCACCCTTATGGTATTTCTTAATCAGCGCTGCAGAAAGATGTGAAA 416
 QY 41 PrometLeuArgInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 Db 417 CCTAGCTACGCGCAGGAGATACAGAGACCTGATTGGAACATTTTGGGTATTAAGCT 476
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
 Db 477 GCTGTTTGGGGTCAACACTGAAATAGCATGCCCAAGAGCTACAGCAGCTGCAGAT 536
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 Db 537 TTCACAGCAAGAGTGGGATGCTGCTCAGAGAGATGAATTCATGACCTGCTCATAA 596
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAspTyrLeu***ThrGlyGln 120
 Db 597 CACATTTGCAAGACTGTGGATTTCACGACGAGTATGATATTATTGTTAACCGGGGACAG 656
 QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
 Db 657 GATTAACCTGTACGATATATGACTTGAACAACTGAGACAGAACCTTAAGGAATTAGT 716
 QY 141 GlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSer 160
 Db 717 GGTCACTACTCTGGATATAAAAAACCTGTGGTGCAGTGAGATTAACAGATCTTTCT 776
 QY 161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeu 180
 Db 777 GCTGATGACAAAACCTGTGACTTTGGATTCATGATGACAGAGATGAAATCTCTA 836
 QY 181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyLysIleLeuValIle 200
 Db 837 AATTTAATATGCTGTGTAGTATGAAATATATCTCGAGGAGAAAGATTTGGTTATA 896
 QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
 Db 897 ACTTATGAGAGATCTATGCTTTTCATAGGCTAGTATTTGGACCAATTAATCTTT 956
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAla 240
 Db 957 GAAGGCTCGCAACATCAATTCGATCTCTCATCCCTGAGAAAGAAATTTCTTGTGCA 1016
 QY 241 GlyGlyGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSer 260

Db 1017 GCGCGTGAAGATTTTAAACTTATATAGTATATATATAGTGAAGAAATAGATCC 1076
 QY 261 TyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeuTyr 280
 Db 1077 TACAAGGACACTTTGGCTTATTCACGTGTGAGATTATCTCGATGAGAACTCTAT 1136
 QY 281 AlaSerGlySerGluAspGlyThrLeuArgLeuTrpGlnThrValAlaGlyLysThrTyr 300
 Db 1137 GCGATGCTCAGAGATGAGACATGAGACTTATGCAAACTGTGTTGAGAAAAAGTAT 1196
 QY 301 GlyLeuTrpLysCysVal**ProGluLysAspSerGlyGluLeuAlaLysProLysIle 320
 Db 1197 GGCCTTGGAAATGTGCTCTTCCTGAAGAGATAGTGTGAGCTGGCAAGCCAAAGATT 1256
 QY 321 GlyPheProGluThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluAsnSerAsp 340
 Db 1257 GGTTCACAGACACAAAGAGAGAGCTA--GAAGAAATTCCTTCAGAGATTCAGAG 1313
 QY 341 SerIleTyrSerSerThrProGluValLysAla 351
 Db 1314 TGCATCTTCTCCTCAGCTCCTGATGTTAAGGCC 1346

RESULT 6
 AR053395 AR053395 1820 bp DNA linear PAT 29-SEP-1999
 LOCUS
 DEFINITION Sequence 2 from patent US 5834240.
 ACCESSION AR053395
 VERSION AR053395.1 GI:5978257
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE
 1 (bases 1 to 1820)
 AUTHORS Bandman O. and Lal P.
 TITLE DNA encoding a transforming growth factor-.beta. receptor
 associated protein
 JOURNAL Patent: US 5834240-A 2 10-NOV-1998;
 FEATURES Location/Qualifiers
 source 1..1820
 /organism="unknown"

BASE COUNT 516 a 370 c 411 g 523 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.73e-153 Length: 1820
 Score: 1779.50 Matches: 337
 Percent Similarity: 97.15% Conservative: 4
 Best Local Similarity: 96.01% Mismatches: 9
 Query Match: 96.71% Indels: 1
 DB: Gaps: 1

US-09-856-836-2 (1-351) x AR053395 (1-1820)

QY 1 MetAlaMetArgInThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
 Db 259 ATGGAATGAGACAGACCGCCGCTACCTGCTTGGACACAGCGGAGCGGTGATTG 318
 QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 Db 319 GCTTCAGTGGATCAACGCTTATGCGTATTTCTTAATCAGCCCTGCAAGATGTGAAA 378
 QY 41 PrometLeuArgInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 Db 379 CCTATGTCAGCAGGAGATATACAGAGACTGATTTGGAACATTTTGGGTCAATTAAGT 438
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
 Db 439 GCTGTTTGGGGTCAACACTGAAATAGAGATGCCAAGACAGCTACAGCAGCTGCAGAT 498
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 Db 499 TTCACAGCAAGAGTGGGATGCTGCTCAGAGAGATGAATTCATGACCCGCTCATATAA 558

QY 101 HistileValIsthrValAspPheThrGlnAspSerAsnTyrLeu***ThrGlyGlyGln 120
 Db 559 CACATTTGCAAGACGTGGATTTCCACCAGAGTACTATTTTGTGAACCGGGGACAG 618
 QY 121 AspIleuLeuAagIleTyrAspIleuAsnLysProGluAgluProlGlyIleSer 140
 Db 619 GATTAACGTGTTACCATATATGACTGAAACCTTAAGCAGAACTTAAGAAATTAGT 678
 QY 141 GlyHisThrSerGlyIleTyrLysAlaLeuTyrPcysSerAspIleuSer 160
 Db 679 GGTCATCTCTGCTATATAAAAGCTGCTGGTGGAGTGAAGTAAACAGATTTCTTCT 738
 QY 161 AlaAspIleuThrValAgluLeuTyrAspHisAlaThrMetThrGluValLysSerLeu 180
 Db 739 GCTGATACAAACCTGTTCCAGCTTGGATCAGCTCTACTATGACAGAACTGAAATCTCTA 798
 QY 181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyIleLeuValIle 200
 Db 799 AATTTTAATGCTCTGTAGTATGATGATATATTTCTTGAAGGAGAGATTTGGTTATA 858
 QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProlIleLysSerPhe 220
 Db 859 ACTATGAGAGATCTATGCTTTTCATAGTCAGTAAGTTGAGCCCAATTAATCTTTT 918
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
 Db 919 GAAGCTCTGCAACCATCAATTCCTGCACTTCCTGCACTGAGAAAGAAATTTCTTGTGCA 978
 QY 241 GlyIleGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSer 260
 Db 979 GGGGGAGAGATTTTAACTTTAAGTATGATATATATGATGAGAGAAATTTAGATCC 1038
 QY 261 TyrIleGlyHisPheGlyProIleHisCysValAlaArgPheSerProAspGlyLeuLeuTyr 280
 Db 1039 TACAAAGGACACTTGTCTCTATCTGCTGAGATTTAGTCTGATGAGAACTCTAT 1098
 QY 281 AlaSerGlySerGluAspGlyThrIleuArgLeuTyrGluThrValGlyIleTyr 300
 Db 1099 GCACAGTGTTCACAAACATGGAACATTTGACATCTGAGCAACCTGTTAGCAAAAGATGAT 1158
 QY 301 GlyLeuTyrLysCysVal***ProGluGluAspSerGlyIleuAlaLysProlGly 320
 Db 1159 GGCTTTGGAATGTGTCTCTGCTGAGAAAGATGAGTGGAGCTGCAAAAGCAAGATT 1218
 QY 321 GlyPheProGluThrAlaGluGluIleuAlaGluGluIleAlaSerGluAsnSerAsp 340
 Db 1219 GGTTCCTGAGAGACACAGAAAGAGCTA--GAAGAAATGCTTCAGAGATTCAGAT 1275
 QY 341 SerIleTyrSerSerThrProGluValLysAla 351
 Db 1276 TGCATCTTTCCTTCAGCTCTGATGTTAAGGCC 1308
 RESULT 7
 LOCUS AB024327 1864 bp mRNA linear PRI 20-DEC-2000
 DEFINITION Homo sapiens pt-wd mRNA for WD-40 repeat protein, complete cds.
 ACCESSION AB024327
 VERSION AB024327.1 GI:4519416
 KEYWORDS pt-wd; WD-40 repeat protein.
 SOURCE Homo sapiens liver cell_line:HepG2 cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Matsuda, S., Katumata, R., Okuda, T., Yamamoto, T., Miyazaki, K.,
 Senga, T., Machida, K., Thant, A.A., Nakatsugawa, S., and Hamaguchi, M.
 TITLE Molecular cloning and characterization of human MAMD, a novel
 protein containing WD-40 repeats frequently overexpressed in breast
 cancer
 JOURNAL Cancer Res. 60 (1), 13-17 (2000)
 MEDLINE 20110763
 REFERENCE 2 (bases 1 to 1864)

AUTHORS Matsuda, S. and Okuda, T.
 TITLE Direct Submission
 JOURNAL Submitted (25-FEB-1999) Satoru Matsuda, Nagoya University School of
 Medicine, Molecular Pathogenesis; 65 Tsurumai-cho Showa-ku, Nagoya
 466-8550, Japan (E-mail:smatsuda@esu.med.nagoya-u.ac.jp,
 Tel:81-52-744-2463, Fax:81-52-744-2464)
 FEATURES
 source
 1. 1864
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="HepG2"
 /tissue_type="liver"
 1..1864
 /gene="pt-wd"
 301..1353
 /gene="pt-wd"
 /function="putative TGF-beta receptor association"
 /note="homologous to TGF-beta-receptor interacting
 protein-1 (TRIP-1)"
 /codon_start=1
 /product="WD-40 repeat protein"
 /protein_id="BA05544.1"
 /db_xref="GI:4519417"
 /translation="MAMRQPLTCSGHTRPVVDLAFSDIPYGFILISACKDQPMR
 OGDGDMIGFTLGHKAVMGATLMDKATKATKATADPTAKMDVSDDELATLHKHT
 VKTVDFQDSNYLTGLGODKRLRYDLNKPAPKEKISGTSIKALMCSDEILS
 ADDKTVRLMDPATMTYKSLFNNMSVSMETIPGELIVITYGKSLAFHSAVSLDPK
 SFEAPATINSASLPBKEFEIVAGDEDRKYVDYNSGESELYKNGHGPVHCVFSD
 GELTASGSDSTLRLMDQVVGKTVGLMKVCVLPEDSGELAKPKIGPETEELEELA
 SENSDCLEPSPDVKA"
 BASE COUNT 528 a 388 c 424 g 524 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.4e-153 Length: 1864
 Score: 1777.50 Matches: 337
 Percent Similarity: 97.15% Conservative: 4
 Best Local Similarity: 96.01% Mismatches: 9
 Query Match: 96.60% Indels: 1
 DB: 9 Gaps: 1
 US-09-856-836-2 (1-351) x AB024327 (1-1864)
 QY 1 MetalMetArgInThrProLeuThrCysSerGlyHisThrArgProValAsp*** 20
 Db 301 ATGGCAATGAGACAGACGCCGCTCACCTGCGCCACAGCCAGCCGTTGATTGG 360
 QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 Db 361 GCGTTGATGACATCAGCCGCTTATGGGATTTCTTAATCAGCGCTGCAAAAGATGTAA 420
 QY 41 PromeIleuArgInGlyAspThrGlyAspTyrIleGlyThrPheLeuGlyHisLysGly 60
 Db 421 CCTATGCTACGCCAGGAGATACAGAGACTGGATTTGAAACATTTTGGCTCATAAAGGT 480
 QY 61 AlaValTyrGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
 Db 481 GCTGTTGGGGTGCACACATCAATTAAGAGTCCACCAAGAGAGCTACAGAGCTGCAGAT 540
 QY 81 PheThrAlaLysValTyrPheAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 Db 541 TTCACAGCCAAAGGTGGAGTGTCTCTCAGAGAGATTAATGATGACCCGCTCATAA 600
 QY 101 HistileValIsthrValAspPheThrGlnAspSerAsnTyrLeu***ThrGlyGlyGln 120
 Db 601 CACATTTGCAAGACGTGGATTTCCACCAGAGTATATTTTGTGAACCGGGGACAG 660
 QY 121 AspIleuLeuAagIleTyrAspIleuAsnLysProGluAgluProlGlyIleSer 140
 Db 661 GATTAACGTGTTACCATATATGACTGAAACCTTAAGCAGAACTTAAGAAATTAGT 720
 QY 141 GlyHisThrSerGlyIleTyrLysAlaLeuTyrPcysSerAspIleuSer 160

[illegible]


```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0447C21
----- Summary Statistics -----
Sequencing vector: M13; 63%
Sequencing vector: Plasmid; 37%
Chemistry: dye-terminator ET; 48% of reads
Chemistry: dye-terminator Big Dye; 52% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181938 bases at least Q40
Consensus quality: 183284 bases at least Q30
Consensus quality: 184429 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 186317; sum-of-contigs
Quality coverage: 6.29 in Q20 bases; sum-of-fp
Quality coverage: 6.62 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.
*
1      4028: contig of 4028 bp in length
4129: gap of unknown length
4129      7967: contig of 3839 bp in length
7968      8067: gap of unknown length
8068      16352: contig of 8284 bp in length
16352      16451: gap of unknown length
16452      26437: contig of 9986 bp in length
26437      35857: gap of unknown length
35858      35888: contig of 9351 bp in length
35889      35988: gap of unknown length

```

FEATURES	*	35989	56349:	contig of 20361 bp in	length
	*	56350	56449:	gap of unknown length	
	*	56450	85530:	contig of 29081 bp in	length
	*	85531	85630:	gap of unknown length	
	*	85631	187017:	contig of 101387 bp in	length.
source			Location/Qualifiers		
			1..187017		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/chromosome="7"		
			/clone="RP11-447C21"		
			1..4028		
misc_feature			/note="assembly_name:Contig26"		
			4129..7967		
misc_feature			/note="assembly_name:Contig27"		
			8068..16351		
misc_feature			/note="assembly_name:Contig28"		
			16452..26437		
			/note="assembly_name:Contig29		
			clone_end:17		
			vector_side:left"		
misc_feature			26538..35888		
			/note="assembly_name:Contig30"		
misc_feature			35989..56349		
			/note="assembly_name:Contig31"		
			56450..85530		
misc_feature			/note="assembly_name:Contig32		
			clone_end:SP6		
			vector_side:right"		
misc_feature			85631..187017		
			/note="assembly_name:Contig33"		
BASE COUNT	53793 a	37172 c	37140 g	58180 t	732 others
ORIGIN					
Alignment Scores:					
Pred. No.:		4.84e-110	length:		187017
Score:		1334.50	Matches:		291
Percent_Similarity:		66.81%	Conservative:		13
Best Local Similarity:		63.96%	Mismatches:		44
Query Match:		72.53%	Indels:		110
DB:		2	Gaps:		3
US-09-856-836-2 (1-351) x AC010101 (1-187017)					
QY	1	MetaMeta.rglinthProleuthrcysserglyhistrharvProvalaIasp***	20		
Db 100587	AVAGCAATGAACACGATCGCTCTACACGTTCTGGCACATATGCGCCGTCGACTTG	100528			
QY	21	AlaPheSerGlyIlethPro-	27		
Db 100527	GCCTTCAGTGGCATTCACTCCCTTAGGGGCTTTTTCCTCCAGACAGCAAACTCTGCTCG	100465			
QY	27		27		
Db 100467	TCGCCACAGCTGAGTGCATGTGTCAATCTTGGCTCACTGCAGCAAACTCCGCTCCAGGT	100408			
QY	27		27		
Db 100407	TCAACTGATTTCTCTGCTCCAGCTCCCGAGTACGTGGATTACAGCGCAGCCTATCAC	100348			
QY	27		27		
Db 100347	TCCTGGCTAATTTTGTATTTTATAGTAGAGAGCGGGGTTTCACCATGTTGGCCACGCTGCT	100288			
QY	27		27		
Db 100287	CTGCAACTTCTGACCTCAGGTGATCCGCTGCTTGGCTCCCAAGTCGTGGGATTACA	100228			
QY	28	-----TTCGTTTTCPhleuIIISerIaIacysIysAs	38		
Db 100227	GGCGTAGACCACTCGCGCTGGCCAGCTTTATATGAGTATTTCTAATACAGTGGCTGCAGAAAG	100168			
QY	38	pGlyIysProMetLeuA.rglnIyAspThrIyAspTPIIeGlyThrPhleuIyH1	58		

```

|||||
Db 100167 CGGCAACCTATGCTATGCGGAGATACAGAGACTGATGGACATTTGGGTCA 100108
|||
Qy 58 slysglyalvaltprglyalathrleuanslysalathrlysalalathrAlaAl 78
|||
Db 100107 TAAAGATGCTGCTGGAATGCAACACTAAATAGATGCGCACAAAGCCAGCACAC 100048
|||
Qy 78 aAlaspherhrAlalyvaltprAlaAlaserglyaspGluLeuMetThrleuAl 98
|||||
Db 100047 TGCAGATTTGACA---AAGTG-TGGATGCTATCCAGGAGATGACCTGCTGCG 99992
|||
Qy 98 ahlsyshsilevallysthrValasphethrglnaspserrAsnTyrLeu**Thrgl 118
|||||
Db 99991 TCATAAACATTAATCAACACATGATTTCAATGAGATAGATTAATTTGTTACCA 99932
|||
Qy 118 ygllynasplysleuLeuarglIetYrAspleuanslyspGluAlgluProlysgl 138
|||||
Db 99931 GGGACAGATTAACCTGTTTGCATATATGACTTGAAACAAACCCGAGGACAACTGAG 99872
|||
Qy 138 vllserglyhsthrSerGlyIlelyslsAlaLeuTprCysSerAspGlnl 158
|||||
Db 99871 AATTGATGATCAGCTGATTAATAAGAGCCCTATGCTACAGTACAGATTAACAGAT 99812
|||
Qy 158 eLeuserAlaaspsdlysthrVal-ArgleuTprAsphsAlathrMetThrGluVal 178
|||||
Db 99811 TCTTTCTGATGACAAACAGTTAAAGCTTGGATCATGCTACTATGACAGAGTGA 99752
|||
Qy 178 ySerleuanspshasmetSerValSerSerMetIuTyrIleProgluGluIlel 198
|||||
Db 99751 AATCTCTAAATTTTATATATGCTCCAGTAGTAAGTAATATATCTGAGCGAGATTT 99692
|||
Qy 198 euVallethrYrGlyAspSerIleAlaPhehlsSerAlaValSerleuGluProl 218
|||||
Db 99691 TGGTAAATTAATGCTGATGCTGATGCTTTTACAGTACAGTATGGAACCAATTA 99632
|||
Qy 218 ySerPhegluaIaProalathrIleasnsSerAlaser**HisProgluGluPhe 238
|||||
Db 99631 AATCTTCAAGCTCCGCAACATTAACCTGATCTTCACTTCACTTCACTTCACTT 99572
|||
Qy 238 euValAlaGlyGlyAspPhehlsLeuTyrIleAspTyrAsnSerGlyGluGlu 258
|||||
Db 99571 TTTGTTGACAGTGTGAAGATGTTAACT-TATAGATATGATGATTAAGTGGAGAAAT 99513
|||
Qy 258 euGluSerYrGlyGlyhlsPheglYProIlehlsCysValArgPheSerProAsp 278
|||||
Db 99512 TGAATCCCTAACAAGACACTGCTGCTATGCTGATGCTGATGCTGATGCTGATG 99453
|||
Qy 278 luleuTyrAlaSerGlySerGlyAspGlyThrleuArgleuTprGlnThrValAl 298
|||||
Db 99452 AACTCTATGCGCAGCTGCTGATGAGATGGAACACTGAGAC--TGGCAAACTGTGTA 99395
|||
Qy 298 yathrYrGlyLeuTprYrCysVal**ProgluGluAspSerGlyGluLeuAla 318
|||||
Db 99394 AATCATATGCGCTTTGGAATGCTGCTTCTTCAAGAGATGCTGCTGCTGCAAGC 99335
|||
Qy 318 rolyslleGlyPheProgluThrAlaGluGluGluAlaGluIleAlaSerGlu 338
|||||
Db 99334 CAAGAATTTGTTTCCAAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 99281
|||
Qy 338 snserAspSerIleTyrSerSerThrProgluVallyAla 351
|||||
Db 99280 AATTCAGATTCATCTATCTTCAACTCTCAAGTTAAGGCC 99240
|||
RESULT 11
AC010132/ 166680 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens BAC clone RP11-111K18 from 7p11.2-p2, complete
sequence.
ACCESSION AC010132
VERSION AC010132.5 GI:10440742
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
9847074
2 (bases 1 to 166680)
Andrews, S., Wohlmann, P., and Scherger, E.
The sequence of Homo sapiens BAC clone RP11-111K18
Unpublished
3 (bases 1 to 166680)
Waterston, R.H.
Direct Submission
Submitted (13-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166680)
Waterston, R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 30, 2000 this sequence version replaced gi:8954217.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH011K18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GWF/CH7/>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-111 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tatenno, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is GS1-308H5, 200 bp overlap. Actual start of this clone is at base position 156485 of GS1-308H5; actual end is at base position 166680 of RP11-111K18.

There are polymorphic base pair differences in the overlap between the clone RP11-111K18 and GS1-308H5.

FEATURES
SOURCE

```
Location/Qualifiers
1..166680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p11.2-p2"
/clone="RP11-111K18"
/clone_lib="RP11"
2..397
/rpt_family="ERV1"
repeat_region
1478..1785
/rpt_family="Alu"
repeat_region
1807..1913
/rpt_family="L1"
repeat_region
2138..2508
/rpt_family="ERV1"
repeat_region
2532..2817
/rpt_family="MALR"
repeat_region
2846..2911
/rpt_family="ERV1"
repeat_region
2912..3043
/rpt_family="Alu"
repeat_region
3044..3115
/rpt_family="ERV1"
repeat_region
3087..3598
/rpt_family="ERV1"
repeat_region
3600..3780
/rpt_family="ERV1"
repeat_region
3781..4090
/rpt_family="Alu"
repeat_region
4091..4257
/rpt_family="ERV1"
repeat_region
4260..4865
/rpt_family="ERV1"
repeat_region
4902..4994
/rpt_family="ERV1"
repeat_region
6077..6610
/rpt_family="ERV1"
repeat_region
7185..7327
/rpt_family="ERV1"
repeat_region
7327..7539
/rpt_family="ERV1"
repeat_region
7579..8074
/rpt_family="ERV1"
repeat_region
8169..9081
/rpt_family="ERV1"
repeat_region
9082..9362
/rpt_family="Alu"
repeat_region
9363..9633
/rpt_family="ERV1"
repeat_region
9634..9943
/rpt_family="Alu"
repeat_region
9944..10150
/rpt_family="ERV1"
repeat_region
10151..10567
/rpt_family="MALR"
repeat_region
10568..12227
/rpt_family="ERV1"
repeat_region
12316..12514
/rpt_family="ERV1"
repeat_region
12532..12582
/rpt_family="ERV1"
repeat_region
13068..13614
/rpt_family="ERV1"
repeat_region
15167..15445
/rpt_family="Alu"
repeat_region
16828..17212
/rpt_family="MALR"
repeat_region
17548..17857
/rpt_family="Alu"
repeat_region
17897..18030
/rpt_family="ERV1"
repeat_region
18365..18652
```

```
repeat_region /rpt_family="Alu"
19118..19746 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
19750..19814 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
19852..20169 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
20224..20412 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
20540..20840 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
20879..20997 /rpt_family="MIR"
repeat_region /rpt_family="MIR"
21649..21848 /rpt_family="MERL_type"
repeat_region /rpt_family="MERL_type"
22158..22177 /rpt_family="MIR"
repeat_region /rpt_family="MIR"
22178..22488 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
22489..22777 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
22778..22919 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
22920..22979 /rpt_family="MIR"
repeat_region /rpt_family="MIR"
23031..23067 /rpt_family="L2"
repeat_region /rpt_family="L2"
23518..23928 /rpt_family="MALR"
repeat_region /rpt_family="MALR"
24091..25047 /rpt_family="L2"
repeat_region /rpt_family="L2"
25094..25312 /rpt_family="L2"
repeat_region /rpt_family="L2"
25375..25549 /rpt_family="L2"
repeat_region /rpt_family="L2"
25663..25817 /rpt_family="MERL_type"
repeat_region /rpt_family="MERL_type"
27098..27212 /rpt_family="L2"
repeat_region /rpt_family="L2"
27657..27925 /rpt_family="MALR"
repeat_region /rpt_family="MALR"
28020..28061 /rpt_family="L2"
repeat_region /rpt_family="L2"
28527..28780 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
29058..29361
```

Alignment Scores:

```
Pred. No.: 1,49e-109 Length: 166680
Score: 1328.50 Matches: 290
Percent Similarity: 66.59% Conservative: 13
Best Local Similarity: 63.74% Mismatches: 45
Query Match: 72.20% Indels: 110
DB: 9 Gaps: 3
```

US-09-856-836-2 (1-351) x AC010132 (1-166680)

```
OY 1 MetAlaMetArgGlnThrProLeuThrcysSerGlyHisThrArgProValAlaSer*** 20
Db 48948 ATGACATGAGACTGATGCTGCTACCTGCTGCGCACATATGCGCCGCTGACTTG 48889
OY 21 AlaPheserGlyIleThrPro----- 27
Db 48888 GCCTTAGTGGCATCATGCCTTAGGGGCTTTTTCCTCCAGACAGACATCTGCTCG 48829
OY 27 ----- 27
Db 48828 TCGCCAGCGCTGAGTGCAGTGTGCATCTGGCTCACTGCACAACTCCGCCAGGT 48769
OY 27 ----- 27
Db 48768 TCAAGTATTTCTCTGCTGAGCTCCCGAGTAGCTGGATTACAGGCGCGCTATAC 48709
```

QY 27 ----- 27
 Db 48708 TCCTGGCTAATTTTGTATTTTGTAGAGACGGGTTCCACCATGTGGCCAGCTGT 48649
 QY 27 ----- 27
 Db 48648 CTCGAACCTTCGACCTGAGGTGATCCGCTGCTGCCCAAGAGTGGGATTTACA 48589
 QY 28 ----- TyGlyTyrPheLeuIleSerAlaCysIysAs 38
 Db 48588 GGGCTGAGGCACTGGCGCTGGCCAGCTTATAGTATTTCTTAATCATGCTCTGCAAGA 48529
 QY 38 PGLIYSPrometLeuArgInglIysPthrgIyAspTPRIegIYThPheLeuGlyH 58
 Db 48528 CGGCAAACTATGCTATGTCAGGAGAGATACAGGAGACTGGATTGGAACAATTTGGGTCA 48469
 QY 58 SLVSGIYAlaValITrpGlyValIethrLeuAnlysAspAlaIethrIysAlaIethrAlaI 78
 Db 48468 TAAAGATGCTGTGGAATGCAACACTGAAATGAGATGCCACCAAGACACCCACAGCAGC 48409
 QY 78 AAlaAspPheThrAlaIyValITrpAspAlaValISerGIyAspGIuLeuMetThrLeuAl 98
 Db 48408 TGGAGATTTACACA---AAGTG-TGGAGTGTATCCAGAGAGATGATGACCTGGC 48353
 QY 98 AHISLYSHISIIeValIlyThrValAspPheThrGIAspSerAsnTyrLeu**ThrgI 118
 Db 48352 TCATAAACACATATTCAGACTATGATTCATGACAGAGATGATTAATTTGTTAACAG 48293
 QY 118 YGIYGIAspIyAspIyLeuArgIleTyrAspIyAspIyProGIuAlaGIuProIySGI 138
 Db 48292 GGGACGAATTAATCTTTTGGCTATATGACTGAAACAAACCCGAAAGCACACCTGGGGA 48233
 QY 138 IILeserGIyHisThrSerGIyIleIySlyAlaLeuTyrCysSerAspIySGInII 158
 Db 48232 AATTAGTGCATCAGCTGATGATTAATAAAGGCCATGAGTACAGAGATTAACAGAT 48173
 QY 158 ELEUSERAlaAspAspIySthrVal-ArgIleutrpAspHisAlaITrmMetThrgIuVal 178
 Db 48172 TCTTTTCTTGAGAGCAAAACCTGTAGACTTGGATCTGCTACTATACAGAAAGTGA 48113
 QY 178 YSERLeuAnpPheAnMetSerValSerSerMetGIuTyrIleProGIuGIuIleI 198
 Db 48112 AATCTTAATATTTTATATGTTGGCTAGTAGATGATGAAATATTTCGAGAGAGAGTTT 48053
 QY 198 EUValIleThrTyrGIyArgSerIleAlaPheHisSerAlaValISerLeuGIuProIle 218
 Db 48052 TGGTAAATACCTATGCTGATCCAGTGCCTTTGACAGTACAGTAAGTTTGAACAATTA 47993
 QY 218 YSERPheGIuAlaIAspAlaIethrIleAsnSerIAser**HisProGIuIySGIuPhe 238
 Db 47992 AATCTTTCAAGCTCTGCAACATTAACCTGAAATCTTTCACCTTGAGAAAGAAATTTTC 47933
 QY 238 EUValAlaGIyGIyGIuAspPheIySthrIySthrIyAspTyrAsnSerGIyGIuIle 258
 Db 47932 TTGTTTCAGGTGTGTAAGATGTAAACT-TATTAAGTATGATGATTAATAGTGAAGAAT 47874
 QY 258 EUGIUSERTyrIySGIyHisPheGIyProIleHisCysValaIArgPheSerProAspGIyG 278
 Db 47873 TGAATCTCAACAAAGACACTGCTGCTCTATTCACCTGTGAGATTTGATCTGATGAG 47814
 QY 278 ILEUUTyrAlaSerGIySerGIuAspGIyThrIleuArgLeuTyrPgiInthValValGIyL 298
 Db 47813 AACTTATGCAAGTGTCTGGAAGATGGAACACTGAGAC--TGGCAAACTGTGGTATGAA 47756
 QY 298 YSThrTyrGIyLeuTyrIyCysVal**ProGIuGIuAspSerGIyGIuLeuAlaIySP 318
 Db 47755 AATCATATGCGCTTTTGAAGATGTGCTTCTGAAAGAGTACTGTGACACTGGCAAGC 47696
 QY 318 ROLYSIIeGIyPheProGIuThrAlaGIuGIuIleuAlaGIuGIuIleAlaSerGIuA 338
 Db 47695 CAAGAGTTGGTTTCCAAAGACAAAGAGAGAG-----CTAGAAGAAATTTGCTCAGAGA 47642
 QY 338 snSerAspSerIleTyrSerSerThrProGIuValIySAla 351

Db 47641 ATTGAGATTCATCTATTTCTTCAACTCTTCAGATTAAGGCC 47601
 RESULT 12
 AC020884_3/c
 WPCOMMENT
 Sequence split into 4 fragments
 Fragment Name Begin End
 AC020884_0 1 110000
 AC020884_1 100001 210000
 AC020884_2 200001 310000
 AC020884_3 300001 373038
 Continuation (4 of 4) of AC020884 from base 300001 (AC020884 Mus musculus clone RP23-
 Alignment Scores:
 Pred. No.: 1, 12e-91 Length: 73038
 Score: 1128.00 Matches: 263
 Percent Similarity: 80.29% Conservative: 14
 Best Local Similarity: 76.23% Mismatches: 52
 Query Match: 61.30% Indels: 22
 DB: 2 Gaps: 5
 US-09-856-836-2 (1-351) x AC020884_3 (1-73038)
 QY 1 MetalameIarGlnThrProLeuThrCysSerGIyHisThrArgProValIAsp** 20
 Db 25489 GTGGCCATGAGGAGACTGCTCCTCCTCAGGCGGCGGCTGGTGGGATTTG 25430
 QY 21 AlaPheSerGIyIleThrProIyGIyTyrPheLeuIleSerAlaCysIyAspGIyLys 40
 Db 25429 TCCCTCAGCAGCAATCCCGCTAAAGGCTGCTTGTGATCAGTGTGCAAGATGGCCAA 25370
 QY 41 PrometLeuArgInglIyAspThrgIyAspTPRIegIYThr-PheLeuGIyHisIySGI 60
 Db 25369 CTAATGCTCCACACAGGAGATACAGAGACTAGATTGGAACATTTTGGGTATAAAG 25310
 QY 60 YAlaValITrpGlyAlaIethrIleu---AsnIyAspAlaIethrIysAlaIethrAlaI 79
 Db 25309 TGGTGTGGGGTCAACAATGACTTAATAAGATGCCCAACAACTGCTACAGCTGC 25250
 QY 79 AspPheThrAlaIyValITrpAspAlaValISerGIyAspGIuLeuMetThrLeuAlaI 99
 Db 25249 AGACATACAC-----GTGATGATGCTGTCTCAGGAGATGAAATGATACCTGGCTCA 25196
 QY 99 SLVSHISIIeValIySthrValAspPheThrGIAspSerAsnTyrLeu**ThrgIyGI 119
 Db 25195 TCAGCAACATGTCAAGACTGTGAAGTTCAACAGATAGACAGTACTGCTGACTGGCA 25136
 QY 119 YGIAspIyLeuLeuArgIleTyrAspIyAspIyProGIuAlaGIuProIySGIuI 139
 Db 25135 -CAGGATTAACCTGTGACATACAGACTGGA-AAACCTGAAAGCAACCTTAAGAAAT 25078
 QY 139 ESerGIyHisThrSerGIyIleIySlyAlaLeuTyrCysSerAspIySGInIleI 159
 Db 25077 TAATGCTCACACTTCCGCTATTAAGAGCTGTGTGTCAGAGTAAATATAGTTCT 25018
 QY 159 USERAlaAspAspIySthrValArgIleutrpAspHisAlaIethrMetThrgIuValIyS 179
 Db 25017 TTCAGCTGATGATTAACCTGTTGGCTGCTGCAATCACACAGTGCACAAAGTGAATC 24958
 QY 179 ILEAspPheAnMetSerValSerSerMetGIuTyrIleProGIuGIuIleIyLeuVal 199
 Db 24957 TTAATAATGT-AATATGCTTTT-AGCAGCACAAATATATATTCCTGAGGAGATCTTGGT 24900
 QY 199 ILEThrTyrGIyArgSerIleAlaPheHisSerAlaValSerLeu-----GIuPT 216
 Db 24899 T-----ATGCTTTTCAATGTCAGATTAATTAATTTGTCGGAGGCC 24858
 QY 216 OILEYSerPheGIuAlaIAspAlaIethrIleAsnSerIAser**HisProGIuIySGI 236
 Db 24857 AATTAATCTTGTGAAGCTCTGCGACCATCAATTTGTCATCTCTCATCAGAGAGAGA 24798
 QY 236 U-PheLeuValAlaGIyGIyGIuAspPheIySthrIySthrIyAspTyrAspIyS 256

RESULT	13
AL645802/c	
LOCUS	226250 bp DNA linear ROD 05-APR-2002
DEFINITION	Mouse DNA sequence from clone RP23-128D9 on chromosome 11, complete sequence.
ACCESSION	AF645802
VERSION	AL645802.11 GI:19572035
KEYWORDS	HMG.
SOURCE	house mouse.
ORGANISM	Mus musculus

COMMENT

numquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced by 119335832.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the R6C1-23 Mouse PAC library
constructed by the group of Pleter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
RECTOR: p3ACE3.6.

FEATURES

source

BASE COUNT	64093 a	48521 c	50155 g	63751 t
ORIGIN				

US-09-856-836-2 (1-351) x AL645802 (1-226520),

QY 21 AlaPhseerGlyIleThrProTyrGlyTyrrpheLeulleSerAlaCysLysaspGlyLys 40
 :::::::::: ||| ||| ||| |:::::::::::::::::::
Db 142383 TCCTTCAGCAGCATCCGCCCTAAAGCGTGCTTTGTGATCAAGTGGCTTGCAAGAAGTGGCAA 142322

41 FLOWCCLGAGLGLNYASPRITGLYASPRITGLYTHR-PHEUGLYHISLYSGI 60
 |||||
 Db 142323 CCTATGCTCCACCAGGAGATACAGAGACTAGATTGSAACATTTTGGGTCATAAAG 14226

Db 142263 TGCTGTTGGGTTACAACATTCAGCTAATAAGATGCCACCAAACTGCTACAGACGCTGC 14220

```

Db 142203 AGACATCACA-----GTGTGATGATCCTGTCACAGAGATGATGATGACCTGCTCA 142155
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 99 slyshisil[eval]lvsThrVal[Asp]pethrc|nasSerAseentvrt[ou****h-ct]oci 110

```

Db 142149 TCAGCACATTGTCAGAGACTGTGAAGTTCACACAGGATAGCAGTTACCTGCTGACTGGGCA 142090

DB	142089	-CAGGATTAACACTGTTGCACATACACGACTTGA- AAACCTGAGCAGAACCTAAGGAAT	1420332
QY	139	eSerGlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspLysGlnIleLe	159

QY 159 uSerAlaAspAspIysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSe 179
|||||
|||||

OY 179 rLeuaSnPhaeasnetSerValSerSermetGluTrpIleProGlutylcIueLauva 199
| |||||
b 141911 TTTAAATTCCT- AATTGCACCGCCGGT- ACCAGCAGCAAAAATAAATTAATGAATGTTTGTTGTGTTT

QY 199 LLEThrTYGLyArGSerIleAlaPheHisSerAlaValSerIeu-----GluPr 216
 Db 141853 T-----ATTGGCTTTTCAATGTGCAGACGATATTTATTAATTTGTCGCGAGGC 141912

[illegible]

DB 141751 ACCTTTTTCGTGCAGCGTGAGAACC-TTTAACTGTAATAAGTACTTCTATATAGGGTGAC 141693

Ddb 141692 AACAAATTGAATCCTACAAAGGTCAACATTTGGTTCATTCAGTGCG - AGATTGAGTCCG 141635

Ddb 141634 ATGGTGAACCTATGCCACTGGTGCTGTAAAGATGGGACACCGAATTGTGGCAAACTGTGG 141575

296 a[G]vlystThrtyglVleutTrlvscysval***Proctlnclywsscdccatcgtt... 315

```

Db 1415174 TAGAANAACCTATGCGCTGTGGAAATGTCGACTTCCAGGAGAAGACAGTGSGGAACATCGG 141515
OY      316   lalysProlYSilleglyPheProglutrhAlaglUglUleuAlaglUtlleAlas 336
          |||||||.....|||
Db 141514 CAAGACCAAAAGATGTGATTCTCAGAAACAGCAGAAAG-----CTGAGAACAAATAATT 141461
OY      336   ercluaSenSer 339
          |||||...:::
Db 141460 CAGAGAAAAGCT 141450

RESULT 14
LOCUS    ACI09605
DEFINITION Mus musculus chromosome 11 clone rp22-470o15 strain 129S6/SVEVTac, complete sequence.
ACCESSION ACI09605
VERSION   ACI09605.11 GI:21327406
KEYWORDS  HTG.
SOURCE    house mouse.
ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE AUTHORS Tian,R., Zhang,C., Restrepo,D. and Roe,B.A.
            TITLE   Direct Submission
            JOURNAL  Unpublished
            REFERENCE 2. (bases 1 to 170393)
            AUTHORS   Tian,R., Zhang,C., Restrepo,D. and Roe,B.A.
            TITLE     Direct Submission
            JOURNAL   Submitted (06-FEB-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
            REFERENCE 3. (bases 1 to 170393)
            AUTHORS   Tian,R., Zhang,C., Restrepo,D. and Roe,B.A.
            TITLE     Direct Submission
            JOURNAL   Submitted (21-MAY-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
            REFERENCE 4. (bases 1 to 170393)
            AUTHORS   Tian,R., Zhang,C., Restrepo,D. and Roe,B.A.
            TITLE     Direct Submission
            JOURNAL   Submitted (04-JUN-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
            REFERENCE 5. (bases 1 to 170393)
            AUTHORS   Tian,R., Zhang,C., Restrepo,D. and Roe,B.A.
            TITLE     Direct Submission
            JOURNAL   Submitted (06-JUN-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
            COMMENT   On Jun 6, 2002 this sequence version replaced gi:21321794.

FEATURES             Source
source               Location/Qualifiers
                    1..170393
                        /organism="Mus musculus"
                        /strain="129S6/SVEVTac"
                        /db_xref="taxon:10090"
                        /chromosome="11"
                        /clone="rp22-470o15"
                        /clone_1lb="RPC1 - 22 (129S6/SVEVTac) Mouse BAC Library"
BASE COUNT      48092 a 37599 c 36934 g 47768 t

ORIGIN
Alignment Scores:
Pred. No.:       3..23e-85      Length:      170393
Score:           1062.00        Matches:      258
Percent Similarity: 79.01%      Conservative: 13
Best Local Similarity: 75.22%    Mismatches:   56
Query Match:     57.72%         Indels:       23

```

DB:	10	Gaps:	5
US-09-856-836-2 (1-351) x AC109605 (1-170393)			
Qy 1 Metalamela	rglnthrProleuthrCysserglnhsthrAlpProvalAlasp***	20	
Db 100944	GTGCGCATGAGGCGAGACTGCTCACTTCTCCAGGCCCAT -GCCGCTGGTGGAATTTG	101002	
Qy 21 Alapheser	gyllethrProtyrglyThrPheleuileSerAlacysAspLylys	40	
Db 101003	TCCTTCAGACATCCCGCCCTAAAGCTGCTTCTATAGTGGCTTGCAAAAGATGGCCA	101062	
Qy 41 Prometleu	arglnglyAsprrhglyAsprrlgelyThr - Pheleuglyhsls	60	
Db 101063	CCCTATGCTCCACGAGGAGATACAGGAGCTAGGTGGAACTTTTGGGTCATTAAGG	101122	
Qy 60 yAlaVal	trpGlyAlaThrleu - AsnlyAspAlaThrlyAlaAlaThrAlaAl	79	
Db 101123	TGCTGGTTGGGGGCAACATTTGATTAAGATGGCCAAACATGCTACAGAGCTGC	101182	
Qy 79 asPhe	thrAlaValtrpAspAlaValserGlyAspGlyleuMetthrleuAla	99	
Db 101183	AGACATCA - - - - - GGGGTGATGCTGCTCAGGAGATGATGATGACCTGGCTCA	101236	
Qy 99 slyhsl	ileVallysthrValasPhehrhrglnAspserAsnTytleu***rhGlyG	119	
Db 101237	TCAGCACTTGTCAAGCTGGAAGTTCACACAGGATGAGATGACTCTGTGAGGCA	101296	
Qy 119 yGln	AsprrlyleuAlaArglyleuAsnlyProGlyAlaGlyProlysglyll	139	
Db 101297	-CAGATTAACATACACA - - - - - CTTGAAAAACCTGAAGCAGACCTTAAGAAAT	101346	
Qy 139 eSer	glnhsthrsergyllelyslsAlaLeuTrpCysserAspAsprrlysglnlele	159	
Db 101347	TGATCTCACTCTCCGGTATTAAGAGCTCT - TGCTGAGTGAAGATTAATAGATCT	101405	
Qy 159 uSer	AlaAsprrlysthrValArgleuTrpAspAlaThrMetthrGlyValysse	179	
Db 101406	TTCCAGCGGATGATTAACCTGTTGGCTCTGGCATTACACAGATGCAAGATGAATC	101465	
Qy 179 rleu	AsnPhesMetSerValSerMetGlyThrleuProGlyGlylylleuVal	199	
Db 101466	TAAAAATG - AATATGCTTT - AGCAGCACAATAATATTCCTGAGGACAAGATCTGGT	101523	
Qy 199 lile	thrTyrrGlyArgSerileAlaPhehslserAlaValserleuGlyProilelys	219	
Db 101524	TATTACTTTCGGGACTTGGTAT - - - - - TAAAGCCGAGACCAATTAAT	101568	
Qy 219 ePhe	glnAlaProAlaThrileAsnSerAlaser***hslProGlyGlyleu - Pheleu	238	
Db 101569	CCTTTGAACCTCTCGACATCAATTTGCTATCTTCATGTAGAAAGAAATTTTTTT	101628	
Qy 239 Val	AlaGlyGlyAlaAspPhehslserlysthrTyrrAsprrlyAsnserGlyGlyGlyleu	258	
Db 101629	GTTCGACGGTGGAAAGC - TTTAACCTGTAATGTAATTTTAATGGTCGAGAAATTA	101687	
Qy 259 Glu	SerTyrrlyGlyhslPhehslProleuhsCysValaArgPheSerProAspGlyGly	278	
Db 101688	GAACTCTAACAAATGACACTTGGCTCATTCAGTGTG - AGATTCAGTCTGATGGTGA	101745	
Qy 279 Leu	TyrAlaSerGlySerGlyAspGlyThrleuArgleuThrProGlyThrValGlylys	298	
Db 101746	CTCTATGCCAGTGTTCGAATATGGACACGAAATTTGTGGCAACTGTGGTAGGAAG	101805	
Qy 299 Thr	TyrGlyleuTrpLysCysVal**ProGlyGlyAspSerGlyGlyleuAlaValPro	318	
Db 101806	ACCTATGGCTGTGGAATGTACTTCTCTAGAGAAACAATGGGGAACTGGCAAAACA	101865	
Qy 319 lye	lileGlyPheProGlyThrAlaGlyGlyleuAlaGlyGlylleuAlaSerGlyAsn	338	
Db 101866	AAGATCTGATTTCCAGAAACACAGAGAG - - - - - CTGGAAGAAATTAATTCAGAGAAA	101919	
Qy 339 Ser	339		

```

Db 101920 gcc 101922
RESULT 15
LOCUS AY061000
DEFINITION Drosophila melanogaster HL01517 full length cDNA.
ACCESSION AY061000
VERSION AY061000.1 GI:16768657
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy. Presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
FEATURES
source
1..1591
location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y, cn bwsp"
/db_xref="taxon:7227"
/map="59E3-59E3"
1..1591
/gene="CG3957"
/note="alignment with genomic scaffold AB003461"
/db_xref="FLYBASE:FBgn0034876"
345..1331
/gene="CG3957"
/note="longest ORF"
/codon_start=1
/product="HL01517p"
/protein_id="AAL28548.1"
/db_xref="GI:16768658"
/db_xref="FLYBASE:FBgn0034876"
/translating="MAANNLRQIPDTCSGHTRPVVHLDFSDICDAGYFLISACKDGSB
MLRIGDGDWGTGFEHKGAVNATLNRNATLASGADFGKVMNVTGAEHRSOP
KHIVSAFVDSSENIYTGSEKLVRFNLEOPAOEERAGHGAIKRALFCGDKC
IIISAEKTVRLMRMTGIEYORLOPNSNPSTLSSDNTHTTISHSSTISFEWIDL
KKLKEVPTNVASLSHPDKHYVCGGDFKMTKFDYITGNETESKGRHGPVHSVK
FSPDGLVASSSEGTIRLMTQVIGKTYGLMKCTEPADLGNSSISPREVAVAN"
BASE COUNT
436 a 451 c 425 g 279 t
Alignment Scores:
Pred. No.: 9.18e-87 Length: 1591

```

```

Score: 1052.50
Percent Similarity: 69.07%
Best Local Similarity: 53.87%
Query Match: 57.20%
DB: 3 Gaps: 3
US-09-856-836-2 (1-351) x AY061000 (1-1591)
QY 3 MetArgLthrProLeuThrCysSerGlyHisThrArgProValAlaLysPhe***AlaPhe 22
Db 360 TTCGAGCAAAATCCCGTGGACCTTCAGTCCACACCGCCGCTGCTCCACCTGCACCTTC 419
QY 23 SerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLysProMet 42
Db 420 AGCGACATCTCGACACCCCGCTTCTTCATCTCGGCTGCAAGATGCGACGCGATG 479
QY 43 LeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaVal 62
Db 480 CTCGGCAGCGGCGACACCGCGCTGTTGGGACCTTCGAGGAGACACAGGCGCGCTG 539
QY 63 TrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThr 82
Db 540 TGGACGCCACGCTGACACCGAAATGCCACCGCGCTCGGAGCGCGCTTCACCTCACC 599
QY 83 AlaLysValIleTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLysHisIle 102
Db 600 GCGACGCTGGAATGCGGTGACCGCGCGCGGCGGCGGCGCTCGGAGCGCGCTTCAC 659
QY 103 ValLysThrValAlaAspPheThrGlnAspSerAsnTyrLeu***ThrGlyGlyLysAsp 122
Db 660 GTGAAGAGGCTGCTTCACGCGGATTCGAGACATAGTACTGCGACGACGACGACGAG 719
QY 123 LeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSerGlyHis 142
Db 720 CTGTGCGCGCTTCACTGACGACGACCGCGGCGGCGGCGGAGGAGGAGTACCGCGCAT 779
QY 143 ThrSerGlyIleLysAlaLeuThrPcysSerAspLysGluIleLeuSer---Ala 161
Db 780 ACGGGCGCATCAACGCGCGCTTCTGCGCGCGCGACAGTGTGATCATCTCGCGCGCC 839
QY 162 AspAspLysThrValArgLeuThrPaspHisAlaThrMetThrGluValLysSerLeuAsn 181
Db 840 GAGGACAGACGCTGCGCTGCGGCGCGATGACGAGGCGGCGGCGGCGGCGCGCTGCG 899
QY 182 PheAsnMetSerValSerSerMetGluTyrIleProGluGlyLysIleLeuValIleThr 201
Db 900 TTCACAGCAATCCGACAGCGCTGAGATCTCCAGCGCACACACATCACTACCATATCG 959
QY 202 TyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGlu 221
Db 960 CACGGCTCTCTGATTAAGCTTCTGGAGATCGACGCGTGAAGAGCTGAAGAGCGTGAAA 1019
QY 222 AlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluLysPheLeuValAlaGly 241
Db 1020 GTGCGGACGACGAGCGGCGCTGCGGCGCATGTTGCGATCGGATTAAGCATGTGTTGTG 1079
QY 242 GlyLysAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluLysLeuGluSerTyr 261
Db 1080 GCGCAGCAGCTTAAAGATGTACAAATTCGACATCAACAGAAAGCAAAATTAATCTTTC 1139
QY 262 LysGlyHisPheGlyProIleHisCysValAlaArgPheSerProAspGlyLysLeuTyrAla 281
Db 1140 AAGGACATTTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199
QY 282 SerGlySerGluAspGlyThrLeuArgLeuThrProGluThrValAlaGlyLysThrTyrGly 301
Db 1200 AGCGGCTCCGAGGACGCGACGCTCGGCTGTGGAGACACACCGTGGCAAGACCTAGGCG 1259
QY 302 LeuTrpLysCysVal***ProGluGluAspSerGlyLysLeuAlaLysProLys----- 319
Db 1260 TTGTGGAAGTGCACGAGCGCGCGGACCTGCGCAATTCATCACTGCTGCGCGCGGAGG 1319
QY 319 ----- 319

```


Db 1320 CAGCAMAAGTGTGTCGCGCAGCAGAGTCGCTGTACACAGCCACGCTGCACCGCG 1379
 Oy 320 -----TLeGlypheProGluThrAlaGluGluLeuAlaGluGlu--- 333
 Db 1380 ACGCAACGAAAGGGCATGTGATTGAGAGCTACGGCTGAGATCCGATTATCAGAACAGAA 1439
 Oy 334 -----TLeAlaSerGluAsnSerAspSerIleTyrSerSer 345
 Db 1440 CAACTCGTAGTGGGGCTAC-AATAACGATTCTTCATATTCACAC 1483

Search completed: February 6, 2003, 23:30:17
 Job time : 3659 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 6, 2003, 13:06:48 ; Search time 2160 Seconds

(without alignments)
2631.767 Million cell updates/sec

Title: US-09-856-836-2
Perfect score: 1840
Sequence: 1 MAHQPLTSCGHRPVYDX.....EELASSENSDSYSTPEVKA 351

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09856836/runat.31012003.144238.22208/app.query.fasta.1.519
-DB=EST -OPMT=fastp -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-USER=US09856836.ecgn.1.1.763.ernat.31012003.144238.22208 -NCPU=6 -ICPU=3
-NO_XLRY -NO_MMP -IARECOVERY -NRG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YCAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST : *
1: em_estba : *
2: em_esthu : *
3: em_estin : *
4: em_estnu : *
5: em_estov : *
6: em_estpl : *
7: em_estro : *
8: em_hic : *
9: gb_estl : *
10: gb_estl2 : *
11: gb_hic : *
12: gb_estl3 : *
13: gb_estl4 : *
14: gb_estl5 : *
15: em_estfun : *
16: em_estom : *
17: gb_gss : *
18: em_gss_hum : *
19: em_gss_inv : *
20: em_gss_pln : *
21: em_gss_vrt : *
22: em_gss_fun : *
23: em_gss_mam : *
24: em_gss_mus : *
25: em_gss_other : *
26: em_gss_pro : *
27: em_gss_rod : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1784.5	97.0	1860	11	AF161496	AF161496 Homo sapi
2	1422	77.3	904	9	AU123736	AU123736 AU123736
3	1380.5	75.0	1174	14	BM802585	BM802585 AGENCOURT
4	1325	72.0	944	9	AL517952	AL517952 AL517952
5	1313.5	71.4	1055	13	BM472016	BM472016 AGENCOURT
6	1302.5	70.8	880	14	BQ219033	BQ219033 AGENCOURT
7	1267	68.9	902	14	BQ889506	BQ889506 AGENCOURT
8	1265	68.8	875	14	BQ439153	BQ439153 AGENCOURT
9	1235	67.1	757	12	BG687473	BG687473 AGENCOURT
10	1228.5	66.8	822	9	AJ395256	AJ395256 AJ395256
11	1217.5	66.2	1019	13	BM450309	BM450309 AGENCOURT
12	1203	65.4	800	9	AU123939	AU123939 AGENCOURT
13	1193	64.8	796	12	BG778828	BG778828 AGENCOURT
14	1184	64.3	901	9	AL556515	AL556515 AL556515
15	1182	64.2	1086	14	BQ433088	BQ433088 AGENCOURT
16	1179	64.1	901	9	AL518570	AL518570 AL518570
17	1176.5	63.9	879	14	BQ222878	BQ222878 AGENCOURT
18	1173	63.7	881	9	AU132581	AU132581 AU132581
19	1150	62.5	1038	9	AL545602	AL545602 AL545602
20	1141.5	62.0	704	12	BG740119	BG740119 AGENCOURT
21	1139	61.9	885	14	BQ883467	BQ883467 AGENCOURT
22	1130.5	61.4	874	9	AL570903	AL570903 AGENCOURT
23	1121	60.9	748	13	B1090693	B1090693 AGENCOURT
24	1115	60.6	951	9	AL551950	AL551950 AL551950
25	1115	60.6	959	9	AL547128	AL547128 AL547128
26	1113.5	60.5	833	9	AU119614	AU119614 AU119614
27	1108	60.2	962	9	AL539156	AL539156 AL539156
28	1106	60.1	967	12	BM164469	BM164469 AGENCOURT
29	1105	60.1	952	9	AL539237	AL539237 AL539237
30	1100	59.8	973	9	AL547220	AL547220 AL547220
31	1099	59.7	776	10	BM725392	BM725392 AGENCOURT
32	1096.5	59.6	685	14	BM742526	BM742526 AGENCOURT
33	1096.5	59.6	685	14	BQ672095	BQ672095 AGENCOURT
34	1095	59.5	955	14	B1712640	B1712640 AGENCOURT
35	1092.5	59.4	638	13	B1088605	B1088605 AGENCOURT
36	1088	59.1	664	13	B1596603	B1596603 AGENCOURT
37	1084	58.9	956	9	AL566260	AL566260 AGENCOURT
38	1079	58.6	656	13	B1112632	B1112632 AGENCOURT
39	1074	58.4	675	10	BE258152	BE258152 AGENCOURT
40	1071.5	58.2	861	13	B1550693	B1550693 AGENCOURT
41	1066	57.9	633	13	B1792000	B1792000 AGENCOURT
42	1065	57.9	904	12	BE684268	BE684268 AGENCOURT
43	1058.5	57.5	705	10	BE544830	BE544830 AGENCOURT
44	1057.5	57.5	684	10	AV723717	AV723717 AGENCOURT
45	1056.5	57.4	1041	13	BM563365	BM563365 AGENCOURT

ALIGNMENTS

RESULT 1	AF161496	1860 bp	MRNA	linear	HTC 22-MAY-2001
LOCUS	AF161496				
DEFINITION	Homo sapiens HSPC147 mRNA, complete cds.				
ACCESSION	AF161496				
VERSION	AF161496.1				
KEYWORDS	GI:684515				
SOURCE	HTC				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 1860)				
AUTHORS	Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G., Shen, Y., Fan, H.Y., Lu, G., Zhong, M., Xu, X.R., Han, Z.G., Zhang, J.W., Tao, J., Huang, Q.H., Zhou, J., Hu, G.X., Gu, J., Chen, S.J. and Chen, Z.				

TITLE Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells

JOURNAL Genome Res. 10 (10), 1546-1560 (2000)

MEDLINE 20499367

PUBMED 11042152

REFERENCE 2 (bases 1 to 1860)

AUTHORS Ye, M., Zhang, Q. H., Zhou, J., Shen, Y., Wu, X. Y., Guan, Z. Q., Wang, L., Fan, H. Y., Mao, Y. F., Dai, M., Huang, Q. H., Chen, S. J. and Chen, Z.

TITLE Human full length cDNA cloned from cd34+ stem cells

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1860)

AUTHORS Ye, M., Zhang, Q. H., Zhou, J., Shen, Y., Wu, X. Y., Guan, Z. Q., Wang, L., Fan, H. Y., Mao, Y. F., Dai, M., Huang, Q. H., Chen, S. J. and Chen, Z.

TITLE Direct Submission

JOURNAL Submitted (23-JUN-1999) Shanghai Institute of Hematology, Shanghai Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II, Shanghai 200025, P. R. China

FEATURES Location/Qualifiers

SOURCE 1. 1860

ORGANISM "Homo sapiens"

DB_XREF "taxon:9606"

CLONE "CBFBN08"

CELL_TYPE "cd34+ stem cells"

TISSUE_TYPE "blood"

CDS 302..1354

codon_start 1

product "HSPC147"

protein_id "AA029111.1"

db_xref "GI:6841516"

translation "MANRQTPLESCGHTRPVVDLAFSGITPYGYFLISACKDKPMIR /GTGDMWIGTFLGKGAAGATLNDKATKATAAFTAAVWAVSDELMTLAKHI /KTVVPTQDSNYLTGGODKLRIDYDNKPEAPEKETSIGIKKALMSDEKQILS /ADDTVRLMDATMTETVEKSLNPNMSVSMYIPGELIVTGYSLFHSVSLDPJK /SEEPATINSLSPPEKEPLVAGCEDEPKLYKYVNSGEELSEVKGHGPIHCVRFSD /EELVASEDETLKQTVGKTYGKTYGLKCVLPEDSGLAKPKRIGPFPETTELLEELA /SENDCIEPSPDYKA"

BASE COUNT 521 a 390 c 425 g 524 t

ORIGIN

Alignment Scores:

Pred. No.: 8.5e-214 **Length:** 1860

Score: 1784.50 **Matches:** 338

Percent Similarity: 97.44% **Conservative:** 4

Best Local Similarity: 96.30% **Mismatches:** 8

Query Match: 96.98% **Indels:** 1

DB: 11 **Gaps:** 1

US-09-856-836-2 (1-351) x AF161496 (1-1860)

QY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaasp** 20

DB 302 ATGGCAATGAGACAGACGGCGCTGCTGCGCACACGACCGCGTGGTGTATTTG 361

QY 21 AlaPheSerGlyTleThrProTyrGlyTyrPheLeuIleSerAlaCysIleAspIleLys 40

DB 362 GCGTTCAGTGCATACGCGCTTATGGGTAATTCATACGCGCTGCAAGATGCTAA 421

QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisIleSgIy 60

DB 422 CCTATGCTACGCCAGGAGATACAGAGACTGATTTGGAACTTTTGGGTCTATAAGT 481

QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaasp 80

DB 482 GCGTTTGGGGTGACACACTGATAGATGCCACCAAGCAGTACAGACCTGCAGAT 541

QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100

DB 542 TTCACAGCCAAAGTGTGGATGCTGTCAGAGATGAATTAAGCCCTGCTCATAA 601

QY 101 HisIleValIleThrValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyGln 120

DB 602 CACATTGTCAAGCTGTGATTTACGACGAGATAGTAATTAATTATTGTAACCGGGGACAG 661

QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140

DB 662 GATTAACCTTTACGATATATGACTTGAACCAACCTGAAGCAGAACTAAGGAATTAAGT 721

QY 141 GlyHisThrSerGlyTleLysLysAlaLeuTyrCysSerAspAspLysGluIleLeuSer 160

DB 722 GGCATACCTCTCGATATAAAAAGCTGTGTGTCAGTGAAGATTAACAGATTCTTCT 781

QY 161 AlaAspAspLysThrValArgLeuThrPaspHisAlaThrMetThrGluValLysSerLeu 180

DB 782 GCGATGACAAAACCTTGCTGACTTGGATCATGCTACTATGACAGAACTGCTCA 841

QY 181 AsnPhaAsnMetSerValSerSerMetGluTyrIleProGluGlyGluLeuValIle 200

DB 842 AATTTAATATGCTGTGTAGTATGATGATATATCCCGAGAGAGAGATTTTGGTTTAA 901

QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220

DB 902 ACTTATGAGAGATCATGCTTCTTCAATGAGTCAATGATTTGAGCCCAATTAATCTTT 961

QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAla 240

DB 962 GAAGCTCCCTGCAACATCATCTGATCTTCACTGATCGAGAAAGAAATTTCTGTGCA 1021

QY 241 GlyGlyGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSer 260

DB 1022 GCGGCTCAAGATTTTAACTTTAAGTATGATATATGATGAGAGAAATTAAGATCC 1081

QY 261 TyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyLeuLeuTyr 280

DB 1082 TACAGAGGACACTTGTGCTTATTCATCTGCTGATGATTTAGTCTGATGAGAACTAT 1141

QY 281 AlaSerGlySerGluAspGlyThrLeuArgLeuTrpGlnThrValAlaGlyLysThrTyr 300

DB 1142 GCAAGTGTTCAGAAATGAGACATTGACACTATGCCAAACTGTGGTAGAAAGAGTAT 1201

QY 301 GlyLeuTyrPlyScyVal**ProGluGluAspSerGlyGluLeuAlaLysProLysIle 320

DB 1202 GGCCTTGGAAATGTGCTTCTGAGAAAGATGATGCTGAGCAAGCCAAAGATTT 1261

QY 321 GlyPheProGluThrAlaGluGluGluGluLeuAlaGluIleAlaSerGluAsnSerAsp 340

DB 1262 GGTTCACAGACACAGAAAGAGAGACTA--GAAGAAATTCCTCAGAGATTCAGAT 1318

QY 341 SerIleTyrSerSerThrProGluValLysAla 351

DB 1319 TGCATCTTCTCCTCAGCTCCTGATGTTAAGGCC 1351

RESULT 2

LOCUS A0123736 904 bp mRNA linear EST 01-AUG-2002

DEFINITION A0123736 NT2RM2 Homo sapiens cDNA clone NT2RM2000935 5', mRNA

ACCESSION A0123736

VERSION A0123736.1 GI:10948452

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 904)

AUTHORS Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and Isogai, T.

TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S., Isogai, T.)

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975
Fax: 81-438-52-3986

Email: genomics@hri.co.jp
HRI human cDNA project: 5'-6' end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source

Location/Qualifiers

1..904

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="NT2RM200935"

/clone_lib="NT2RM2"

/cell_type="teratocarcinoma"

/cell_line="NT2"

/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

BASE COUNT 257 a 173 c 214 g 256 t 4 others

ORIGIN

Alignment Scores:

Pred. No.: 1.54e-168 Length: 904
Score: 1422.00 Matches: 279
Percent Similarity: 95.59% Conservative: 3
Best Local Similarity: 94.58% Mismatches: 11
Query Match: 77.28% Indels: 3
DB: 9 Gaps: 0

US-09-856-836-2 (1-351) x AU123736 (1-904)

1 MetAlaMetArgGlnThrProLeuThCysSerGlyHisThrArgProValValAsp*** 20
24 ATGCATATGAGACAGACGCGCTACCTGCTGGCCACGACCGCGCTGGTATTG 83
21 AlaPheSerGlyLeuThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
84 GCCCTTCAGTGCACGACGCTTATGGTATTCTTATTCAGCCCTTCAAGATGTTAA 143
41 ProMetLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
144 CTTATGCTACGCGGAGGAGTACGAGAGCTGATGGAACATTTTGGGTCAATAAGT 203
61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaAlaAlaAsp 80
204 GCTGTTGGGGTGAACACGATTAAGATGACCAACCACTACAGAGCTGGAGAT 263
81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
264 TTCACAGCCAAAGTGGGATGCTGCTCAGAGATGAATGATGACCCGCTCATATA 323
101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu***ThrGlyGlyGln 120
324 CACATGTCACAGCTGATGATTCACGCAAGATAGTAATATTGTTAACCGGGGAGAG 383
121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
384 GATTAACCTGTCGATATATGACTTGAACAAACCTGAAGACGAACCTTAAGAAATTA 443
141 GlyHisThrSerGlyLysLysAlaLeuTyrPyrSerAspAspLysGluIleLeuSer 160
444 GGTCACTCTCTGGTATTAATAAAAGCTGCTGTCAGAGAGATTAACAGATTCCTTCT 503
161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeu 180
504 GCGATGACAAAACCTGCTGATTTGGGATCATGCTACATATACAGAAAGTGAATCTCTA 563
181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyGluIleLeuValIle 200
564 AATTTTAATATCTCTGTTAGTATGATGAAATATATCTCGAGAGAGAGATTTGGTTATA 623
201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSerPhe 220
624 ACTTATGGACGATCTATCTCTTCTATAGTGCAGTAAGTTGGACCAATTAATATCCTTT 683

221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
684 GAACTCTCTGCAACCAATTCATCTCTCATCTCTGCTGGAATATATTTCTTGGTGA 743
241 GlyGlyGluAspPheLysLeuTyrLysTyrAspPyrAsnSerGlyGluLeuGluGlu 260
744 NGCGGTGAACATTTTAACCTTATTAAGTATGATTAATTAAGTGAACAAATTAAGATC 803
260 TTYrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeu 280
804 CTC-AAGGACACTTGGTCTCTATATCATCTGCGTGAAGATTTAATCTGATGAGAACTCTA 862
280 rAlaSerGlySerGluAsp-GlyThrLeuArgLeuTrpGln 293
863 TGCCAGNGGTTCANMAAGATTGGAACATTGGGACTNCGCA 903

RESULT 3
BM802585
LOCUS
DEFINITION
AGENCY: NIH-MGC
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM802585 1174 bp mRNA linear EST 05-MAR-2002
5', mRNA sequence.
BM802585
BM802585.1 GI:19119408
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1174)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNI)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNI at:
http://image.llnl.gov
Plate: LAM12324 row: b column: 24
High quality sequence stop: 656.

FEATURES
source
Location/Qualifiers
1..1174
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5574839"
/clone_lib="NIH-MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: PCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."

BASE COUNT 369 a 232 c 257 g 316 t

ORIGIN

Alignment Scores:
Pred. No.: 4.29e-163 Length: 1174
Score: 1380.50 Matches: 288
Percent Similarity: 88.46% Conservative: 11
Best Local Similarity: 85.21% Mismatches: 27
Query Match: 75.03% Indels: 12
DB: 14 Gaps: 6

US-09-856-836-2 (1-351) x BM802585 (1-1174)

22 PheSerGlyLeuThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLysPro 41
3 TTCAGTGCACACGCGCTTATGGTATTCTTATATCAAGCGCTTGCAGAAAGATGAACCT 62

QY 42 MetLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAla 61
 Db 63 ATGCTACGGCAGGAGATACAGAGACTGATGGACATTTTGGGTCAATAAGAGTCT 122
 QY 62 ValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaThrAlaAlaIleAspPhe 81
 Db 123 GTTGGGCTGCAACACAGATGAAGATGCCAACAGCAGCTACAGCAGCTGCGATTTTC 182
 QY 82 ThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLysHis 101
 Db 183 ACACGCCAAATGTGGGATGCTGCTCTCAGAGATGAATGTAGACCTGGCTCATAAACAC 242
 QY 102 IleValLysThrValAspPheThrGlnAspSerAsnTrpLeu***ThrGlyGlyLys 121
 Db 243 ATGTCTAAGACTGTGATTTCCAGCAGATAGTAAATATTTGTTAACCGGGGAGAGAT 302
 QY 122 LysLeuLeuArgTrpIleTrpAspLeuAsnLysProGluAlaGluProLysGluIleSerGly 141
 Db 303 AAAGCTTTAGCATATATGACTTGAACAACTGACAGACAGAACTTAAGAAATAGTGT 362
 QY 142 HisThrSerGlyIleLysLysAlaLeuTrpCysSerAspLysGlnIleLeuSerAla 161
 Db 363 CATACTCTGATATAAAAAAGCTCTGTGTCAGTCAGATGAACAGATTTCTTCTGCT 422
 QY 162 AspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeuAsn 181
 Db 423 GATACCAAACTGTTCGACTTGGATCATCTTACTATGACAGAAAGTAAATCTCTAAT 482
 QY 182 PheAsnMetSerValSerSerMetGluTrpIleProGluGlyGluIleLeuValIleThr 201
 Db 483 TTTATATAGCTGTAGTATAGTATGAAATATATCTCTGAGGAGAAATTTGGTTATACT 542
 QY 202 TyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGlu 221
 Db 543 TATGACATCTATTTGCTTTTCATAGTCAGTAACTTGGACCCCAATTAATCTTTGAA 602
 QY 222 AlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAlaGly 241
 Db 603 GCTCTGACAACTCAATTCGATCTCTCATCTTGAAAGAAATTTCTTCTGACAGC 662
 QY 242 GlyLysAspPheLysLeuTrpLysTrpAspTrpAsnSerGlyGluLeuGluSerTrp 261
 Db 663 GGTGAAGATTATTAACCTTATATAGTATATATATAGTGAAGAAATTAATCTCTTAC 722
 QY 262 LysGlyHisPheGlyProIleHisCys-ValArgPheSer-ProAspGlyGluLeu-Tyr 280
 Db 723 AAGGACACTTGTGCTATATCAGCTGGTGAGATTTAGTCCCGATGAGAACATCTTAT 782
 QY 281 AlaSerGly---SerGluAspGly-ThrLeuArgLeuTrpGlnThrValVal---GlyLys 298
 Db 783 GCCATGTGGTTTCAGAAAGATGGAACATTTGAGACTATGCAAAACTGTGGTGAAGAAA 842
 QY 298 sThrTrpGlyLeuTrpLysCysVal***ProGluGluAsp-----SerGlyGluLeu 315
 Db 843 AACGATGTGGCTTTGAAATATGTCCTTCCGCAAAAAAAGTGGGTGGACCTGC 902
 QY 315 uAlaLysProLysIleGlyPhe---ProGluThrAlaGlu-GluGluLeuAlaGluLys 334
 Db 903 CACAACCCCAAAAGATTTGGTTCCTCCGCAAAACCAACCAAGAGGACACCTTTAAAGAAA 962
 QY 334 lAlaLysSerGluAsn---SerAspSerIleTrpSerSerThrPro 347
 Db 963 TTGCTTCCAAAAAATTTGCAATGAGCATCTTTCTTCCAGCT 1006
 RESULT 4
 AL517952 944 bp mRNA linear EST 13-FEB-2001
 LOCUS AL517952 LTI_NFLO11.NBC1 Homo sapiens cDNA clone CS0DA004YH09 5
 DEFINITION AL517952 LTI_NFLO11.NBC1 Homo sapiens cDNA clone CS0DA004YH09 5
 prime, mRNA sequence.
 ACCESSION AL517952
 VERSION AL517952.1 GI:12781445
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 944)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..944
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DA004YH09"
 /clone_id="LTI_NFLO11.NBC1"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 259 a 197 c 235 g 250 t 3 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 3 02e-156 Length: 944
 Score: 1325.00 Matches: 256
 Percent Similarity: 98.10% Conservative: 2
 Best Local Similarity: 97.34% Mismatches: 5
 Query Match: 92.01% Indels: 2
 DB: 9 Gaps: 0
 US-09-856-836-2 (1-351) x AL517952 (1-944)
 QY 1 MetAlaMetArgGlnGlyAspThrProLeuTrpCysSerGlyHisThrArgProValAlaAsp*** 20
 Db 158 ATGGCAATGACAGACAGCGCGCTACCTGCTGCGCACAGCGACCCGCTGGTGAATTGG 217
 QY 21 AlaPheSerGlyTrpLeuProGlyTrpPheLeuIleSerAlaCysLysAspGlyLys 40
 Db 218 GCCCTGAGGCACTACGCCCTTATGGGATTTCTTATATACAGCGCTTGCAAAAGATGTTAA 277
 QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 Db 278 CCATGCTACGCCAGGAGATACAGAGACTGATTTGAAACATTTTGGGTCAATAAGCT 337
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
 Db 338 GCTGTGGGCTGCAACACTGATTAAGATGACCAACAGCAGCTACAGCAGCTGAGAT 397
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 Db 398 TTACAGCCCAAAAGTGGATGCTGCTCAGAGATGAATGTAGTACCGGCTCATAAA 457
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTrpLeu***ThrGlyGlyGln 120
 Db 458 CACATTTGCAAGACTGTGATTTCCAGCAGATAGTAAATATTTGTTAACCGGGGAGACAG 517
 QY 121 AspLysLeuLeuArgTrpIleTrpAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
 Db 518 GATTAACCTGTACGATATATGACTTGAACAAACCTGAAACAGAACTTAAGAAATTAGT 577

QY 141 GlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSer 160
 Db 578 GGTATCTCTGTTGTAATAAAAAAGCTGTGTGAGATGAAGATTAACAGATCTTCT 637
 QY 161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetTrpGluValLysSerLeu 180
 Db 638 GCTGATACAAAAGCTGTTCCACTTGGGATCATGCTACTATGACAGAAAGAAATCTCTA 697
 QY 181 AspPheAsnMetSerValSerSerMetGluTrpIleProGluGluIleLeuValIle 200
 Db 698 AATTATATATGCTCTTACTAGATGAGATATATTTCTTCTGAGAGAGATTTGGTTATA 757
 QY 201 ThrTyrglyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
 Db 758 ACTATGAGACGATCTATGCTTTTCATAGTCAGTAAGTTGGACCAATTAATCTCTT 817
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAla 240
 Db 818 GAAGCTCTGCAACATCAATCTGCATCTTCATCTGAGAAAGAAATTTCTTGTGCA 877
 QY 241 GlyGlyGluAspPheLysLeuTrpLysTyrglyAspTrpAsnSerGlyLysGluLeuGluSer 260
 Db 878 GSGGGTGAAATTTTAACTTATAGTATGATTATATAGTGGGAGRA-TTAGAATC- 935
 QY 261 TyrlYsgly 263
 Db 936 TACAAGGG 944
 RESULT 5
 BM472016 1055 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT 6464560 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5581667
 DEFINITION 5', mRNA sequence.
 ACCESSION BM472016
 VERSION BM472016.1 GI:18521058
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1055)
 AUTHORS NIH-MGC http://mgi.nhl.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.lnl.gov
 Plate: LMNL2341 row: 0 column: 12
 High quality sequence start: 3
 High quality sequence stop: 658.
 Location/Qualifiers
 source
 1. 1055
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5581667"
 /clone_id="NIH_MGC_67"
 /lssue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORE6; Site:1: Notti;
 Site:2: Sali; Cloned unidirectionally. Primer: Oligo
 Average insert size 1.75 kb. Library constructed by Life
 Technologies."
 BASE COUNT 308 a 222 c 247 g 277 t 1 others
 ORIGIN
 Alignment Scores: 1.04e-154 Length: 1055
 Pred. No.:

Score: 1313.50 Matches: 268
 Percent Similarity: 87.74% Conservative: 11
 Best Local Similarity: 84.28% Mismatches: 29
 Query Match: 71.39% Indels: 10
 DB: 13 Gaps: 4
 US-09-856-836-2 (1-351) x BM472016 (1-1055)
 QY 1 MetalMeThrArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp** 20
 Db 68 ATGGCAATGAGACAGACGCGCTCACTCTCTGGCACACGACCGGTGTTATTTG 127
 QY 21 AlaPheSerGlyIleThrProTrpGlyTrpPheLeuIleSerAlaCysLysAspLys 40
 Db 128 GCTTACAGGATGACAGCGCTTATGGGTATTTCTTAATCAGCGCTTGCAAGATGTGATA 187
 QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 Db 188 CCTATGCTACGCGCAGGAGATACAGAGACAGATTTGGATGACATTTTGGTCATTAAGT 247
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaTrpLysAlaAlaThrAlaAlaAsp 80
 Db 248 GCTGTTGGGGTCAACACTGAAATAGATGACCAAAAGAGCTACAGACAGCTCAGAT 307
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 Db 308 TTCACAGCCAAAGTGTGGATCTCTCTCAGAGATGATGATGATGATGATGATGATGAT 367
 QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTrpLeu**ThrGlyGlyGln 120
 Db 368 CACATTTGCAACAGCTGTGATTTCCAGCAGATAGTATATTGTTGTTAACCGGGGAGAC 427
 QY 121 AspLysLeuLeuArgIleThrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
 Db 428 GATAAAGCTTACGATATATGACTTGAACAAACCTGAAAGCAAGCACTTAAGAAATAGT 487
 QY 141 GlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSer 160
 Db 488 GGTCACTCTGTAATGTAATAAAAGCTGTGTGAGTGAAGATTAACAGATTTCTTCT 547
 QY 161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetTrpGluValLysSerLeu 180
 Db 548 GCTGATGACAAACGTTGCGACTTGGATCATCTCTATGACAGAAAGCAATCTCTA 607
 QY 181 AspPheAsnMetSerValSerSerMetGluTrpIleProGluGluIleLeuValIle 200
 Db 608 AATTATATATGCTCTGTTAGTATGATGATATATTTCTTGGAGAGATTTGGTTATA 667
 QY 201 ThrTyrglyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
 Db 668 ACTATGAGACATCTATGCTTTCATAGTACAGTAGTATGTTGGACCCCATTAATCTT 727
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAla 240
 Db 728 GAACCTCTGCAACATCAATCTGCATCTCTGAGGAGAGATTTGGTTATA 787
 QY 241 GlyGlyGluAsp-PheLysLeuTrpLysTyrglyAspTrpAsnSerGly--GluLeuGlu 259
 Db 788 GCCGCGCAAGATTTTACCTTATAGTATGATTAATATAGGAGAGAAAGATTTTAC 847
 QY 259 uSerTyrglyGlyHis-PheGlyProIleHisCys-ValArgPheSerProAspLys 278
 Db 848 TTCTTACAGGAGACCTTGGGCTTATGATGATTAATATAGGAGAGAAAGATTTTAC 907
 QY 278 u--LeuTyrglyAla-SerGlySerGluAspGlyThrLeuArgLeuTrpGlnThrValVal 296
 Db 908 AAACCTCTATGCCAGTGGTTCACCAAGAGCAACATTGCAACTATGCGCAACTGTGG 967
 QY 297 -----GlyLysThrTyrglyLeuTrpLys--CysVal**Pro 308
 Db 968 AACGCAAAAACGTATGGCTTGGAAATAGTGTGCTCTCT 1011
 RESULT 6

```

BQ219033      880 bp      mRNA      linear      EST 02-MAY-2002
LOCUS         BQ219033
DEFINITION    AGENCOURT_7575739 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6047340
              5', mRNA sequence.
ACCESSION     BQ219033
VERSION       BQ219033.1 GI:20400433
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 880)
AUTHORS       NIH-MGC http://mhc.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgaabs-remail.nih.gov
              Tissue Procurement: ATCC/DCTD/DMP
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              Plate: LLAM13294 row: j column: 13
              High quality sequence stop: 741.
FEATURES
  source
    1..880
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:6047340"
    /clone_lib="NIH_MGC_72"
    /tissue_type="melanotic melanoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: skin; Vector: PCMV-SPORT6; Site: 1: NotI;
    Site: 2: SalI; Cloned unidirectionally. Primer: oligo dT.
    Average insert size 2 kb. Library constructed by Life
    Technologies."
BASE COUNT    249 a 173 c 209 g 249 t
ORIGIN
US-09-856-836-2 (1-351) x BQ219033 (1-880)
Alignment Scores:
Pred. No.:    1..86e-153      Length:      880
Score:        1302.50         Matches:    252
Percent Similarity: 95.93%    Conservative: 7
Best Local Similarity: 93.33%  Mismatches: 9
Query Match:  70.79%         Indels:     2
DB:           14             Gaps:       1
QY 1 MetAlaMetAgtGlnThrProLeuThrcysSerGlyHisThrArgProValValAsp*** 20
Db 1 ATGGCAATGAAGACAGACCGCTGACCTGCTGCGCACAGCAGCCCGCTGATTTG 97
QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
Db 21 GCCCTTCACTGACATCACCCCTTATGCGATTTCTTAATCAGCGCTGCAGAAAGATGTA 157
QY 41 ProMetLeuArgGlnGlyAspThrGlyAspTrpIleGlyTyrPheLeuGlyHisLysGly 60
Db 41 CCTATGCTACGACGAGGAGATACAGAGACTGATGAACTTTTGGTCAATAAGGT 217
QY 61 AlaValTyrPylAlaThrLeuAlaLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
Db 61 GCTGTTGGGGGCAACCTGATATAGATGCGCCACCAAGCAGCTACAGACAGCTCAGAT 277
QY 81 PheThrAlaLysValTyrAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
Db 81 TTCACAGCAAGCTGGATGCTGCTCAGAGATGATGATGATGATGATGATGATGATGAT 337
QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrcLysGlyGln 120
Db 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrcLysGlyGln 120

```

```

Db 338 CACATTGTCAGACAGCTGATTTACAGCAGAGATAGTAATATTATTGTTAAACCGGGGACAG 397
QY 121 AspLysLeuLeuArgLysThrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
Db 121 ATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
QY 141 GlyHisThrSerGlyIleLysLysAlaLeuTyrCysSerAspAspLysGluIleLeuSer 160
Db 141 GGTCACTACTCTGCTGATTAATAAAGCTGTGGGAGAGATGAATTAAGATTTCTTCT 517
QY 161 AlaAspAspLysThrValArgLeuTyrPaspHisAlaThrMetThrGluValLysSerLeu 180
Db 161 GCTGATGACAAACAGCTGCTCAGCTTGGGATCAGTACTATGACAGAGAGAAATCTCTA 577
QY 181 AspPheAsnMetSerValSerSerMetGluTyrIleProGluGluIleLeuValIle 200
Db 181 AATTATTATGCTGCTGATAGTATGAAATATATCTCTGAGGAGAGATTTGGTTATA 637
QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSerPhe 220
Db 201 ACTTATGACAGATCTATGCTTTTCATAGTACAGTACTTGCACCAATTAATCCTTT 697
QY 221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
Db 221 GAACTCTCTGCAACCAATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 757
QY 241 GlyGlyLysAspPheLysLeuTyrLysTyrAspTyrAsnSer-GlyGluGluLeuGluSe 260
Db 241 GGGCGGCAATTTTAACTTATATAGTATGATATATATATATATATATATATATATATAT 817
QY 260 rTyr--LysGlyHisPheGlyProIle 268
Db 260 CTTACCAAGGACCACTTGGTCCCTA 845
RESULT 7
BQ889506      902 bp      mRNA      linear      EST 16-AUG-2002
LOCUS         BQ889506
DEFINITION    AGENCOURT_8119860 Lupskl_dorsal_root_ganglion Homo sapiens cDNA
              clone IMAGE:6178023 5', mRNA sequence.
ACCESSION     BQ889506
VERSION       BQ889506.1 GI:2281520
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 902)
AUTHORS       NIH-MGC http://mhc.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgaabs-remail.nih.gov
              Tissue Procurement: Dr. James R. Lupski
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM13556 row: o column: 16
              High quality sequence stop: 705.
FEATURES
  source
    1..902
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:6178023"
    /clone_lib="Lupskl_dorsal_root_ganglion"
    /sex="male"
    /tissue_type="dorsal root ganglia"
    /dev_stage="adult, 36 yr"
    /lab_host="DH10B"
    /note="Vector: PCMV-SPORT6 (Life Technologies); Site: 1:
    NotI, Site: 2: SalI; cDNA made by oligo-dT priming.

```

Directionally cloned using the following adaptors:

5'-TCGACCCACGCTCCG-3' and
5'-GACTAGTTCAGATCGGAGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupsik, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies.

BASE COUNT 257 a 186 c 223 g 236 t
ORIGIN

Alignment Scores:

Score: 1267.00 Length: 902
Percent Similarity: 95.47% Matches: 248
Best Local Similarity: 93.58% Conservative: 5
Query Match: 68.86% Mismatches: 7
DB: 14 Indels: 5
Gaps: 1

US-09-856-836-2 (1-351) x BQ889506 (1-902)

QY 1 MetAlaMetAArgInThProLeuThCysSerGlyHisThrArgProValAlaAsp*** 20
DB 115 ATGGCAATGAGACAGACGCGCTCAGCTGCTGCGCACACGCGCCGCTGATTG 174
QY 21 AlaPheSerGlyIleThProTyGlyTyPheLeuIleSerAlaCysLysAspGlyLys 40
DB 175 GCCCTTCACTGACATCAGCGCTTATGGTATTTCTTAATCAGCGCTCAAGAGTAA 234
QY 41 PrometLeuArgInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
DB 235 CCTATGCTACGCCAGGAGATACAGAGACTGATGGAACATTTGGGTCAATAAGT 294
QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
DB 295 GCTGTGGGCTGCAACACTGAATAGATGCCACCAACACCTACAGCAGCTGCAGAT 354
QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
DB 355 TTCACACCAAGAGTGGATGCTGCTCAGAGATGATGATGACCTCGCTCATAAA 414
QY 101 HisLeuAlaLysThrValAspPheThrGlnAspSerAsnTyLeu***ThrGlyGlyGln 120
DB 415 CCAATGTCAGACTGGATGATTCACGACAGATAGTAATATTGTTAAACGGGGACAG 474
QY 121 AspLysLeuLeuArgIleTyAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
DB 475 GATTAACGTTCAGCAATATGACTTGACAAACCTGAAGCAGAACCTTAAGAAATTAGT 534
QY 141 GlyHisThrSerGlyLysLysAlaLeuTrpCysSerAspAspLysGlnIleLeuSer 160
DB 535 GGTCAATCTCTGTTAAATAAAAGCTGTGTGAGAGATGAACAGATCTTTCT 594
QY 161 AlaAspAspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeu 180
DB 595 GCGATGACAAAACCTTGACTTGGATCATGCTACTATGACAAAGCTGAATCTCTA 654
QY 181 AsnPheAsnMetSerValSerSerMetGluTyIleProGluGlyLysLeuAlaIle 200
DB 655 AATTAAATATGCTGTAGTATGATGATATATTCCTGAGAGAGACATTTGGTATTA 714
QY 201 ThrTyGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
DB 715 ACTTATGAGCAGATCAATCTTTCTTATGTCAGTAAGTTGGACCAATTAATCCTTT 774
QY 221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
DB 775 CAAGCTCCGCAACCATCAATCTTCATCTTCATCCGAGAAAGAAATTTCTTGCTGA 834
QY 241 Gly--GlyGluAspPheLysLeuTyLysTyL--AspTyAsnSerGlyGluLysGlu 259
DB 835 GCGCGGGGAAATTTTAACTTTATAGTATGATTAATATAGTGCAGAAAGAAATT--- 891

QY 260 SerTyLys 262
DB 892 ---TACAAA 897

RESULT 8
BQ439153
LOCUS

DEFINITION
BQ439153 875 bp mRNA linear EST 24-MAY-2002

AGENCOURT_7827244 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6014262

ACCESSION
BQ439153
VERSION
BQ439153.1 GI:21178229

KEYWORDS
SOURCE
ORGANISM

REFERENCE
1 (bases 1 to 875)

AUTHORS
TITLE
JOURNAL

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
plate: LLNL13208 row: h column: 07
High quality sequence stop: 675.

FEATURES
source
location/Qualifiers

1..875
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6014262"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site: 1: NOT; Site: 2: Salt; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 253 a 175 c 209 g 235 t 3 others
ORIGIN

Alignment Scores:

Score: 1,016.148 Length: 875
Percent Similarity: 1265.00 Matches: 248
Best Local Similarity: 95.80% Conservative: 3
Query Match: 68.75% Mismatches: 8
DB: 14 Indels: 3
Gaps: 0

US-09-856-836-2 (1-351) x BQ439153 (1-875)

QY 1 MetAlaMetAArgInThProLeuThCysSerGlyHisThrArgProValAlaAsp*** 20
DB 90 ATGGCAATGAGACAGACGCGCTCAGCTGCTGCGCACACGCGCCGCTGATTG 149
QY 21 AlaPheSerGlyIleThProTyGlyTyPheLeuIleSerAlaCysLysAspGlyLys 40
DB 150 GCCCTTCACTGACATCAGCGCTTATGGTATTTCTTAATCAGCGCTCAAGAGTAA 209
QY 41 PrometLeuArgInGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
DB 210 CCTATGCTACGCCAGGAGATACAGAGACTGATGGAACATTTGGGTCAATAAGT 269
QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
DB 270 GCTGTGGGCTGCAACACTGAATAGATGCCACCAAGAGAGCTTACAGCAGCTGCAGAT 329

OY 81 PheThAlAlaValTTPaSPaIaValSerGlyAspGluLeuMetThrIleuAlaHisLys 100
 |||||||
 Db 330 TTCACAGCCAAAGTGGAGTGTCTGCAGGAGATTAATGATGACCCGCTGCATATAA 389
 OY 101 HisIleValIysThrValaSPheThGlnAspSerAsnTyrlLeu***ThrGlyGlyGln 120
 |||||||
 Db 390 CACATTCTCAAGACTGCGATTTCACGAGATAGTAATATTGTTAAACCGGGGACAG 449
 OY 121 AspLysLeuLeuAlaGlyIeTyrrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
 |||||||
 Db 450 GATTAACCTGTACCATATATAGACTTGAACAACTGAAAGCAACCTTAAGGAATTAAGT 509
 OY 141 GlyHisThrSerGlyIleLysLysAlaLeuThrPcysSerAspLysGlnIleLeuSer 160
 |||||||
 Db 510 GGTATACCTCTGGATATAAAAGCTCTGTGTCAGTGAAGATTAACATCTTCT 569
 OY 161 AlaAspLysThrValaLarGleuTrpAspHisAlaThrMetThrGluValLysSerLeu 180
 |||||||
 Db 570 GCTGATGCAAAAGCTGTCGACTTGGATCATGCTACTATGACAGAAAGTGAATCTCTA 629
 OY 181 AsnPheAsnMetSerValSerSerMetGluTyrlIleProGluGluIleValIle-200
 |||||||
 Db 630 AATTATATATGCTGTATGATGATGAAATATATCTCTGAAGGAGAGATTTGGTATA 689
 OY 201 ThrTyrglyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSer-Ph 220
 |||||||
 Db 690 ACTATATGACATCATATGCTTTTCATAGTGCATGTAAGTGGACCAATATATACCT 749
 OY 220 egluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAl 240
 |||||||
 Db 750 TGAAGCTCTCCCAACCAATTCATCTTCATCTTCCTGGAAGAAATTCCTTTTTCG 809
 OY 240 aglyGlyGluAspPhe-LysLeuTyrlLysTyrrAspTyrrAsnSerGly-GluGluLeuGlu 259
 |||||||
 Db 810 AGCGGTGMAAGTTTAAACCTTTAAGTATGATTATATAGTGAAGAATAATAGAA 869

RESULT 9 757 bp mRNA linear EST 01-MAY-2001
 LOCUS BG687473 602639445p1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4762475 5'
 DEFINITION mRNA sequence.
 ACCESSION BG687473
 VERSION BG687473.1 GI:13918870
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 757)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rgs@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM616 row: 3 column: 12
 High quality sequence stop: 723.
 Location/Qualifiers
 1..757
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4762475"
 /clone_lib="NIH_MGC_59"
 /tissue_type="muscle/skin"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggc);

Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATAGGC-3' and 3' adaptor sequence: 5'-ATTCTAGACAGCGCGCCGACATG-3'. Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 235 a 120 c 182 g 220 t
 ORIGIN

Alignment Scores:

Pred. No.: 4,83e-145 Length: 757
 Score: 1235.00 Matches: 238
 Percent Similarity: 94.86% Conservative: 2
 Best Local Similarity: 94.07% Mismatches: 12
 Query Match: 67.12% Indels: 1
 Gaps: 0

us-09-856-836-2 (1-351) x BG687473 (1-757)

OY 77 AlaAlaAlaAspPheThAlAlaValTTPaSPaIaValSerGlyAspGluLeuMetThr 96
 |||||||
 Db 1 CGAGCTGACGATTTACAGCCCAAGTGTGGATGCTGTCTCAGGAGATGATGATGAC 60
 OY 97 LeuAlaHisLysHisIleValLysThrValaSPheThGlnAspSerAsnTyrlLeu*** 116
 |||||||
 Db 61 CTGGCTCATAAACACATGTCATGCAAGCTGTCATGACGACGATGATTAATTTGTGA 120
 OY 117 ThrGlyGlyGlnAspLysLeuAlaGlyIeTyrrAspLeuAsnLysProGluAlaGluPro 136
 |||||||
 Db 121 ACCGGGGCAGAGATAAAGTGTACCATATATGCTTGAACAACTGACAGACAACT 180
 OY 137 LysGluIleSerGlyHisThrSerGlyIleLysLysAlaLeuTrpCysSerAspLys 156
 |||||||
 Db 181 AAGGAATTAAGTGCATATCTCTGTTATATAAAAGCTCTGCGTGCAGTGCAGATATA 240
 OY 157 GlnIleLeuSerAlaAspLysThrValaLarGleuTrpAspHisAlaThrMetThrGlu 176
 |||||||
 Db 241 CAGATCTCTTTCGTCGATACAAAGCTGTCGACCTTGGACATCATCTATGACAGAA 300
 OY 177 ValLysSerLeuAsnPheAsnMetSerValSerSerMetGluTyrlIleProGluGlu 196
 |||||||
 Db 301 GTCAATCTCTAATTTATATATGCTGTAGATGATGAAATATATCTCTGAGGAGAG 360
 OY 197 IleLeuValIleThrTyrglyArgSerIleAlaPheHisSerAlaValSerLeuGluPro 216
 |||||||
 Db 361 ATTTGGTTATTAACCTTATGACGATCTATTCCTTTCAATGACAGTAAAGTTGGACCA 420
 OY 217 IleLysSerPheGluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGlu 236
 |||||||
 Db 421 AATTAATCTCTTGAAGCTCTGCAACCATCAATTCGATCTCTTATCTCTGACAGAA 480
 OY 237 PheLeuValaLarGlyGluAspPheLysLeuTyrlLysTyrrAspTyrrAsnSerGlyGlu 256
 |||||||
 Db 481 TTCTCTGTTCAGAGCGGTGAAGATTTAAACTTATTAAGATATATATATATGAGAA 540
 OY 257 GluLeuGluSerTyrlLysGlyHisPheGlyProIleHisLysValaLarPheSerProAsp 276
 |||||||
 Db 541 GAATTGAATCCCTTCAAGGACGACCTTGGTCTTATTCACGCTGTGTGATTAATCTCGAT 600
 OY 277 GlyGluLeuTyrrAlaSerGlySerGluAspGlyThrLeuArgLeuTrpGlnThrVala 296
 |||||||
 Db 601 GGAAGAACTCATGCTGCTGTCAGAAAGATGAAACATTTGACATATGCAAACTGTGTA 660
 OY 297 GlyLysThrTyrglyLeuTyrrLysCysVal***ProGluLysAspSerGlyGluLeu-Al 316
 |||||||
 Db 661 GGAAGAAAGTATGAGCTTGTGACAAATGTGTCTTGMAAGATATGTGTGAGCTGGC 720
 OY 316 alyspProLysIleGlyPheProGluThrAlaGluGlu 328
 ||||| ||| ||||||| |||||||

Db 721 AAGGCAAGATGATTCAGACAGACAGAGAG 757

RESULT 10
AJ395256 822 bp mRNA linear EST 25-JAN-2001

LOCUS AJ395256 dkt426 Gallus gallus cDNA clone 21m101, mRNA sequence.
DEFINITION AJ395256
ACCESSION AJ395256
VERSION AJ395256.1 GI:7125986
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE
1 (bases 1 to 822)
Adrikmanov, I., Lodygin, D., Gerolt, P., Arakawa, H., Law, A., Plachy, J., Korn, B., and Buerstedde, J.M.
A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function
Genome Res. 10 (12), 2062-2069 (2000)

JOURNAL 20568495
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.
Location/Qualifiers

FEATURES
source
1. 822
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="21m101"
/clone_lib="dkt426"
/tissue_type="Bursa of Fabricius"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
BASE COUNT 249 a 150 c 211 g 211 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 3.7e-144 Length: 822
Score: 1228.50 Matches: 238
Percent Similarity: 92.34% Conservative: 15
Best Local Similarity: 86.86% Mismatches: 18
Query Match: 66.77% Indels: 3
DB: 9 Gaps: 1

US-09-856-836-2 (1-351) x AJ395256 (1-822)

QY 52 IleglYThrPhelenglyHistsyglYAlaValTTPglYAlaThrlenuAsnlyspAla 71
|||||
Db 3 ATGGGACGCTTCTAGGTATATAAGTGTCTGCTGGGTGCTACTTTGAAACAAAGATGCC 62
|||||
QY 72 ThrlYsAlaAlaThrlAlaAlaAspPheThrlAlaValTTPAspAlaValsergly 91
|||||
Db 63 ACTAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 122
|||||
QY 92 AspGluLeuMetThrlenuAlaHistsyHstleValysThrlValAspPheThrlasp 111
|||||
Db 123 GATGAACATACATACATGCTGCACAAACATGTCAAAGGTGGATTTTACACAGCAT 182
|||||
QY 112 SerAsnTyrLeu**ThrGlyglYglInaspLyseuLeuArGlyLeTyrAspLeuAsnly 131
|||||
Db 183 AGCAATTAATCTGTAAACAGGTGGACAGATAAATGTTGGCATATGATTTAAAGCAAG 242
|||||
QY 132 ProGluAlaGluProLyseuGluLeuSerGlyHstHstserGlylleYsAlaLeuTTP 151
|||||
Db 243 CCGGAGCAGCAACCTCGATCTGTGTCAGTGGCATCTCTGCTATTAAAGCTTTATGC 302
|||||
QY 152 CysSerAspAspLyseuGluLeuSerAlaAspAspLyseuThrlValArgLeuTTPAspHis 171
|||||
Db 303 AGCAGTGTATGATTAACAGATCTTTGACGCTGATGATTAAGTGTGCCCTCTGGACAGC 362

QY 172 AlaThrMetThrGluVallyssSerLeuAsnPhasmetSerValSermetGluTyr 191
:::|||||
Db 363 AGTACATGACTAGTAAGGAGCAGCTAAATGTTGCTATGCTGTGAGCAGCATGAGTAT 422
|||||
QY 192 IleProGluGlyGluLeuValIleThrTyrGlyArgSerIleAlaPheHstserAla 211
:::|||||
Db 423 GTTCCAGAAAGGCAATTAATCTGTGATACCTATGAGCAAGCATTTCTTTTACAGCGC 482
|||||
QY 212 ValSerLeuGluProIleLyseuPheGluAlaProAlaThrlleAsnSerAlaSer** 231
:::|||||
Db 483 GAAATCTTACAGCAGATTAATCATTTTGAAGCAGCCTGTACATCAATCTGCATCCCT 542
|||||
QY 232 HisProGluLyseuGluProPheLeuValAlaGlyGlyGluAspPheLyseuTyrTyrAsp 251
|||||
Db 543 CACCCAGAGAAAGAAATGTTGGTGTGCAAGGTGGTGAAGATTTAACTGTATAAATATGAC 602
|||||
QY 252 TyrAsnSerGlyGluLeuGluGluSerTyrLyseuHstPheGlyProIleHstCysVal 271
|||||
Db 603 TATAACACAGCAGCAAGCAATCAATCTTCAAAAGGCACTTTGTCATTCACGTGTG 662
|||||
QY 272 ArgPheSerProAspGlyGluLeuTyrAlaSerGlySerGluAspGlyThr-LeuArgLe 291
|||||
Db 663 AGATTAGCCCTGACGCGGAGTATATGCAAGTGTGCTCAGAGATGTACCTTAAGCT 722
|||||
QY 291 UTTPGlnThrlValAlaGlyLyseuThrlTyrGlyLeuTTP-LysCysVal**ProGluGlu 311
|||||
Db 723 GGGGAGCAACAGACAGTAAAGCAATGCTTTGTTGGAGAGTGTACTTCTGTAAGAG 782
|||||
QY 311 sPserGlyGluLeuAlaLyseuProLyseuIleglyPhePro 323
:::|||||
Db 783 AGAATGACAGAA--AGCAGNCACAAAGCCAGGAGACCC 817
|||||

RESULT 11
BM450309 1019 bp mRNA linear EST 05-FEB-2002
LOCUS BM450309
DEFINITION AGNCOURT_6394168 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494620
5', mRNA sequence.
ACCESSION BM450309
VERSION BM450309.1 GI:18499349
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1019)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLE Email: cgaps-remail.nih.gov
JOURNAL Tissue Procurement: ATCC
COMMENT CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LRAM12120 row: d column: 13
High quality sequence stop: 705.
location/Qualifiers
1. 1019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5494620"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10b (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Library constructed by Life Technologies."
BASE COUNT 281 a 215 c 257 g 263 t 3 others
ORIGIN

Alignment Scores:

Pred. No.: 1.33e-142 Length: 1019
 Score: 1217.50 Matches: 250
 Percent Similarity: 89.35% Conservative: 10
 Best Local Similarity: 85.91% Mismatches: 21
 Query Match: 66.17% Indels: 10
 Gaps: 4

US-09-856-836-2 (1-351) x BM450309 (1-1019)

QY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
 DB 161 ATGGCAATGAGACAGACGCCGCTGCTGCTGCGCACACGCCAGCCGCGTGGATTGG 220
 QY 21 AlaPheSerGlyLeuThrProGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 221 GCCTTGATGGCATCAGCGCTTATGGGTATTCTTAATCAGCGCTGCAAGATGTAA 280
 QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 DB 281 CCTATGCTACGCCAGGAGATACAGGAGACTGGATGGAACTTTTGGGTCATTAAGGT 340
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaAlaAlaAsp 80
 DB 341 GCTGTGGGGGTGACACACTGATTAAGGATGCCACAAAGCAGTACAGCAGTGCAGAT 400
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 DB 401 TTCACAGCCAAAGTGGAGTGTCTCTCAGAGAGATGAATGAATGAGACCTGGCTCATAAA 460
 QY 101 HisLeuValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyGln 120
 DB 461 CACATTGTCAGACACTGCGATTTCACGACGATAGTATATTCTTAACCGGGGACAG 520
 QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluLysSer 140
 DB 521 GATTAACCTGATACGATATATGACTTGAACAACTGAAACGAACTTAAGAAATTAAGT 580
 QY 141 GlyHisThrSerGlyIleLysLysAlaLeuThrPyrSerAspAspLysGlnIleLeuSer 160
 DB 581 GGTATATCTGTGTATATAAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 640
 QY 161 AlaAspAspLysThrValArgLeuThrAspHisAlaThrMetThrGluValLysSerLeu 180
 DB 641 GCTGATGCAAAAGCTGTCGACTTGGGATCATCTCTACATGACAGAGTGAATCTCTTA 700
 QY 181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGluGluIleLeuValIle 200
 DB 701 AATTTTATATCTGTGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 760
 QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
 DB 761 ACTTATGACGATCTATATGCTTNTCATAGTGCAGTAAGTTGGACCAATTTAACTCTT 820
 QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGlu-LysGluPheLeuValAl 240
 DB 821 TGAGCTCCGCAACCAATCATCTCTCATCTCTCTGAAAGAAAGAAATTTCTTGTGTC 880
 QY 240 ArgLysGlyLysAspPheLysLeuTyr-LysTyrAspTyrIleAsnSerGlyLysGlu 258
 DB 881 AGCCCGGTGAAAAATTTTAAACTTTTATATAGTATGATTTATTAATGGGGGAGAAAGAT 940
 QY 258 eu-GlusertyrLysGlyHis--PheGlyProIleHisCysValArgPheSerProasp 276
 DB 941 TAAAAATCTTACCAAGGAGAACACTTTGGGGGCC-----TAATTCACCGGGGG 988
 QY 277 GlyGluLeuTyrAlaSerGlySerGlu 285
 DB 989 GGAGAAATTAAGTCCCGGAGATGGAG 1015

RESULT 12
 AUI23939

LOCUS AUI23939 800 bp mRNA linear EST 01-AUG-2002
 DEFINITION AUI23939 NT2RM2 Homo sapiens cDNA clone NT2RM2001349 5', mRNA
 ACCESSION AUI23939
 VERSION AUI23939.1 GI:10948655
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
 Isogai,T.
 HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
 Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
 Y., Sugano,S., Isogai,T.)
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomcs@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 Location/Qualifiers
 1..800
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM2001349"
 /clone_1id="NT2RM2"
 /cell_type="teratocarcinoma"
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
 precursor cells"
 BASE COUNT 227 a 168 c 194 g 211 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.89e-141 Length: 800
 Score: 1203.00 Matches: 233
 Percent Similarity: 98.33% Conservative: 2
 Best Local Similarity: 97.49% Mismatches: 4
 Query Match: 65.38% Indels: 1
 Gaps: 0
 DB: 9
 US-09-856-836-2 (1-351) x AUI23939 (1-800)
 QY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
 DB 80 ATGGCAATGAGACAGACGCCGCTGCTGCTGCGCACACGCCAGCCGCGTGGATTGG 139
 QY 21 AlaPheSerGlyLeuThrProGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
 DB 140 GCCTTGATGGCATCAGCGCTTATGGGTATTCTTAATCAGCGCTGCAAGATGTAA 199
 QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
 DB 200 CCTATGCTACGCCAGGAGATACAGGAGACTGGATTTGGAACATTTTGGGTCAATAAGGT 259
 QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaAlaAlaAsp 80
 DB 260 GCTGTGGGGGTGACACACTGATTAAGGATGCCACAAAGCAGTACAGCAGTGCAGAT 319
 QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
 DB 320 TTCACAGCCAAAGTGGAGTGTCTCTCAGAGATGAATGATGACCTCGGCTCATAAA 379
 QY 101 HisLeuValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyGln 120

```

Db 380 CACATTGCAACACTGATTTTCACGACGAGTAGTAATTTGTTAAACCGGGAGACG 439
QY 121 AsplysleuLeuArgIleTyrAspleuAnlysProgluAlaGluprolysgluIleSer 140
Db 440 GATTAACCTGTCAGCATATATGACTTGAACAACCTGAACGAACTTAAGAAATTAGT 499
QY 141 GLYHSTHrSerGlyIleLysLysAlaLeuTyrCysSerAspAspLysGlnIleLeuSer 160
Db 500 GCTCACTACTGCTGATATAAAAAACCTCTGTGTGACAGTGAAGATTAACAGATTCTTCT 559
QY 161 AlaAspAspLysThrValArgLeuTyrPaspHisAlaThrMetThrGluValLysSerLeu 180
Db 560 GCTGATGACAAACACTGTCGACTTGGCATCATGCTACTATGACGAAAGTGAATCTCTA 619
QY 181 AsnAspAsnMetSerValSerSerMetGluTyrIleProgluLysGluIleLeuValIle 200
Db 620 AATTTAATATGCTGCTAGTAGTATGATGAATATATCTCTGAGGAGACATTTGCTATA 679
QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
Db 680 ACTTATGACCATCTATTGCTTTTCATAGTGCAGTAAGTTCGACCAATTAATCCCTTT 739
QY 221 GluAlaProAlaThrIleAsnSerAlaSer**HisProgluLysGluPheLeuVal 239
Db 740 GAAGCTCCTGC-ACCATCAATTCCTGCATCTCTCATCTGAGAAAGAAATCTTGTGTG 795

RESULT 13
LOCUS BG778828 796 bp mRNA linear EST 15-MAY-2001
DEFINITION 602667790F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:480720 5',
ACCESSION BG778828
VERSION BG778828.1 GI:14049145
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 796)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
http://image.jnl.gov
Plate: LMC1658 row: d column: 23
High quality sequence stop: 717.
Location/Qualifiers
1..796
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="480720"
/clone_lib="NIH_MGC_60"
/tissue_type="adrenocorticaloma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: prostate; Vector: pDR-11B (Clontech);
site_1: SfiI (ggccgcctggcgc); site_2: SfiI (ggccattagcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCCATTTAGCC-3' and 3' adaptor
sequence: 5'-ATCTAGAGCCGAGGCGCGAGATG-dT(30)-BN-3'
(Where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC

```

```

BASE COUNT 245 a 124 c 188 g 239 t
ORIGIN
Alignment Scores:
Pred. No.: 1,07e-139
Score: 1193.00
Percent Similarity: 97.12%
Best Local Similarity: 95.47%
Query Match: 64.84%
DB: 12
gaps: 0
US-09-856-836-2 (1-351) x BG778828 (1-796)

QY 90 SerGlyAspGluMetThrLeuAlaHisLysHisIleValLysThrValAspPheThr 109
Db 2 TCAGAGATGAAATGATGACCTGGCTCATTAACATATGCAAGACTGATTCACG 61
QY 110 GluAspSerAsnTyrLeu***ThrGlyGluAspLysLeuArgIleTyrAspLeu 129
Db 62 CAGGATGATATTTATTTGTTAACCGGGGACAGATTAACCTGTTACGATATATGACTTG 121
QY 130 AsnLysProgluAlaGluProLysGluIleSerGlyHisThrSerGlyIleLysLysAla 149
Db 122 AACAAACCTGAACAGACAACTAAGAAATTAAGTGCATACCTCTGATATAAAAAAGCT 181
QY 150 LeuTyrCysSerAspAspLysGlnIleLeuSerAlaAspAspLysThrValArgLeuTyr 169
Db 182 CTGTGTCAGTAGAGGATTAACAGATCTTCTGTCGATGACAAACCTGTGACTTTGG 241
QY 170 AspHisAlaThrMetThrGluValLysSerLeuAsnMetSerValSerSerMet 189
Db 242 GATCATGCTACTATGACAGAGAAATGCTTAATTTAATATGCTGTAGTATAG 301
QY 190 GluTyrIleProgluLysGluIleLeuValIleThrTyrGlyArgSerIleAlaPheHis 209
Db 302 GAATATATTCCTGAGGAGAGATTTGTTAATTAATGAGCATATTCCTTTTCAT 361
QY 210 SerAlaValSerLeuGluProIleLysSerPheGluAlaProAlaThrIleAsnSerAla 229
Db 362 AGTGCAGTAAAGTTGGACCAATTAATCTTTGAAGCTCTGCAACATCAATTCGCA 421
QY 230 Ser***HisProgluLysGluPheLeuValAlaGlyGlyGluAspPheLysTyrIlys 249
Db 422 TCCTTCATCTGAGAAATAATTTCTTGTGAGGGGGGAGAAATTTAATTAAG 481
QY 250 TyrAspTyrAsnSerGlyLysGluLeuGluSerTyrLysGlyHisPheGlyProIleHis 269
Db 482 TATGATTAATTAATGAGAGAAATTAAGATTCCTCAAGGACACTTGGTCTATTCAC 541
QY 270 CysValArgPheSerProAspGlyLysLeuTyrAlaSerGlySer-GluAspGlyThrLe 289
Db 542 TGTGTGAGATTTAAGTCCATATGAGAACTATGCAAGTGTTCAGAAATGAGAACATT 601
QY 289 LysGluTyrPaspHisThrValAlaGlyLysThrTyrGlyLeu-TripLysCysVal***Proc 309
Db 602 GAGACTATGCGAAACCTGTGTGAGAAACGATATGCTTTGGAATGTGCTTCTCTG 661
QY 309 LysGluAspSer-GlyGluLeuAlaLysProLysIleGlyPheProgluThrAlaGluLys 328
Db 662 AAGAAGATATGCTGCTGACCTGCAACCAAGATTTGTTCCAGAGACAGAAAGAG 721
QY 329 Glu 329
Db 722 GAG 724

RESULT 14
LOCUS AU556515 901 bp mRNA linear EST 16-FEB-2001
DEFINITION AU556515 LTI_NFL006_P12 Homo sapiens cDNA clone CSDBK006Yg21 5
prime, mRNA sequence.
ACCESSION AU556515
VERSION AU556515.1 GI:12899259

```

KEYWORDS EST.
SOURCE human
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 901)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers

FEATURES
source
1. 901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS00K006Y21"
/clone_lib="LTI_NFI006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site-1: NotI; 1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 245 a 191 c 223 g 241 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 1,83e-138 Length: 901
Score: 1184.00 Matches: 241
Percent Similarity: 96.81% Conservative: 2
Best Local Similarity: 96.02% Mismatches: 8
Query Match: 64.35% Indels: 5
Gaps: 0

US-09-856-836-2 (1-351) x AU556515 (1-901)

OY 1 MetalameatarglntthProleuthrcyseryglhsthrArgProvalasp*** 20
Db 153 ATGGCAATGAGACACACCGCGCTCACTGCGGCACACCGCCGCTGATTG 212
OY 21 AlapheserlylterthProtyrlytyrrhelenlleserAlacyslsasplylys 40
Db 213 GCCTCAGCGGATCAGCGCTTATGGTATTTCTTAATCAGCGCTTGAAGATGATAA 272
OY 41 PromettlaurglnglyaspthrlyasprrpjlleglythPhelenglyhstlysgly 60
Db 273 CCTATGCTACGCCAGGACATACAGAGACGTGATTTGGATTAAGT 332
OY 61 AlavaltrpalyalathrleuasnlyasapalathrLysAlaAlaAlaAlaasp 80
Db 333 GGTGTTGGGCGCAACACTGATAGATGCCACCAAGACGCTACACGCTGCAGAT 392
OY 81 PhetrhAlalyValtrpaspAlaValserglyaspgluleumethrleuAlahstlys 100
Db 393 TTCACAGCCCAAGGTGGATGCTGCTCAGGAGATGAATGTGACCCCTGGCTATAA 452
OY 101 HsilevallysthrValasphehrghlnaspserrAsnTyrlleu***Thrglyglylgn 120
Db 453 CACATTGTACAGACTGTGATTTTCACGACGAGATAGTAATATTATTGTAACGGGGGACAG 512
OY 121 AsplysleuAluargllytyraspleuasnlysproglyuAluaglypPolysglulieser 140
Db 513 GATAACTGTTCAGCATATATGACTTGAACAAACCTGAAGCAACCTAAGGAATTAAGT 572
OY 141 GlyhsthsrserglyllylelyslsAlaleuTrpcysersAspaspysgluleuSer 160

Db 573 GGTCACTTCTGTGATATAAAAAAGACTCTGTGTCAGATGAGATTAACAGATCTTCT 632
OY 161 AlaspaspyslysthrValArgleutTrpasphtsAlarhmetThrglyuValysSerleu 180
Db 633 GCTATGACAAAGCTGTTCGACTTTGGATCATCTACTATGACAGAGTAAATCTCTA 692
OY 181 AsnPhasmetSerValserSerMetglutrylleprogluglyluilleleuValille 200
Db 693 AATTATATATCTCTGTAGATAGATATATATCTCTGAGGAGACATTTGGTATTA 752
OY 201 ThrTyrglyargerlleAlaPhethtserAlaValserleuGluprolielysSerhe 220
Db 753 ACTTATGACCATCTATCTTTCATAGTCAGTAAGTTGGACCAATTAATTC-TTT 811
OY 221 GluAlaprcAlathrIleasnSerAlaser***HisprogluLyglupheleuValaAla 240
Db 812 GAGCTCTCTGACACATC-AATTTCATCTCTTCATCTCTGAGAA-GAATTTCTTGTCA 869
OY 241 GlylyglyuaspPhelyleuTyrllystyrasp 251
Db 870 GCGGTGATAT-TTAAAT-T-TATAGTATGAT 900

RESULT 15
LOCUS BQ433088
DEFINITION BQ433088 1086 bp mRNA linear EST 24-MAY-2002
ACCESSION AGENCOURT_7766619 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015268
VERSION BQ433088.1 GI:21172164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1086)
NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL13211 row: b column: 05
High quality sequence stop: 655.
Location/Qualifiers
1. 1086
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6015268"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site-1: NotI; Site-2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 281 a 266 c 254 g 283 t 2 others

ORIGIN

Alignment Scores:
Pred. No.: 4.55e-138 Length: 1086
Score: 1182.00 Matches: 237
Percent Similarity: 91.39% Conservative: 7
Best Local Similarity: 88.76% Mismatches: 21
Query Match: 64.24% Indels: 4
Gaps: 0

US-09-856-836-2 (1-351) x BQ433088 (1-1086)

```
OY      1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
      292 ATGGCAATGAGACAGACGCCGCTCAGCTGCTGGCCACACGCCGCTGGTTGATTTG 351
OY      21 AlaPheSerGlyIleThrProTyrGlyIlePheLeuIleSerAlaCysLysAspGlyLys 40
      352 GCCTTCAGTGGCATCAGCGCTTATGSGTATTTCTTAATCAGCGCTTGCAGAAAGATGTAAA 411
OY      41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
      412 CCTATGCTACGCCAGGAGATACAGAGACTGATTCGACACTTTTGGCTCATTAAGGT 471
OY      61 AlaValIlePheGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
      472 GCTGTTGGGGTCAACACTGTAATAGATGCCACCAAGCAGCTACAGACGCTCAGAT 531
OY      81 PheThrAlaLysValIleTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
      532 TTCACAGCCCAAGTGGGATGCTCTCAGAGATGATGATGATGATGATGATGATGATGAT 591
OY      101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyGln 120
      592 CACATTTGCAAGACTGTGATTTCCAGCAGATAGTAATTAATTTGTTAACCGGGGACAG 651
OY      121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
      652 GATTAACCTGTTACGATTAATGACTGTAACCAACCTGAACGAGAACCTTAAGCAATTAAGT 711
OY      141 GlyHisThrSerGlyIleLysAlaLeuThrProCysSerAspLysGlnIleLeuSer 160
      712 GGTCACTACTTGGTATTAATAAAGCTCTGTGTCAGTGGAGATTAACAGATTCCTTCT 771
OY      161 AlaAspAspLysThrValArgLeuThrAspHisAlaThrMetThrGluValLysSerLeu 180
      772 GCTGATGACAAACTGTTCGACTTGGATCATCTACTATGACAGAAAGTGAATCTCTA 831
OY      181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyIleLeuValIle 200
      832 AATTTTAATATGCTCTTAANAGNATGGATATATTCCTGAGGAGAGAT-TTGGTTATA 890
OY      201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
      891 ACTTATGGACGATCTAATGCTTTCT-AATGCCCTAAGTTTGACCCATTAAATCCCTT 949
OY      221 GluAlaProAlaThrIleAsnSerAlaSer**HisProGlu-LysGluPheLeuValAl 240
      950 GAAGGTCTCTGAACCAATCTGATCCCTCATCTGAAAAAATAATTTCTGTTGC 1009
OY      240 aglyGlyLysAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGlu-GluLeuGln 260
      1010 ACGGGGGGAAAATTTTAACCTTTAATAGTTGATTAATAAAGGGGAAAAAATAATAAA 1069
OY      260 eTyrLysGlyHisPhe 265
      1070 TCTACAGGGAACCTTT 1086
Db
```

Search completed: February 7, 2003, 00:00:04
Job time : 2182 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 6, 2003, 13:03:57; Search time 57 Seconds

(without alignments)
1888.483 Million cell updates/sec

Title: US-09-856-836-2

Perfect score: 1840

Sequence: 1 MAMQPTLTCGHRFPVVDX.....EELASNSDSITSTPEVKA 351

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USFTO.spool/US09856836/runatc.31012003_144237_22183/app.query.fasta_1.519
-DB=Issued_Patents_NA -OFMT=fastaf -SUFFIX=tbl -MINMATCH=0.1 -LOOPEXT=0
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=tbl -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09856836.eccl 1.1.17 -gunatc.31012003_144237_22183 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEOUTERY -NMG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
6: /cgn2_6/prodata/1/ina/5B.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1779.5	96.7	1820	2	US-08-828-922-2
2	425	23.1	686	4	US-09-040-984-36
3	425	23.1	686	4	US-09-123-912-36
4	425	23.1	686	4	US-09-643-597-36
5	279	15.2	5152	4	US-09-690-364-10
6	274	14.9	3747	4	US-09-690-364-17
7	268.5	14.6	7042	4	US-09-092-508-1
8	268.5	14.6	7042	4	US-09-435-115-1
9	268.5	14.6	7042	4	US-09-098-310-1
10	268.5	14.6	7042	4	US-09-690-364-21
11	268.5	14.6	7075	4	US-09-092-508-15
12	268.5	14.6	7075	4	US-09-435-115-15

13	238.5	13.0	2152	1	US-08-188-582-17	Sequence 17, Appl
14	238.5	13.0	2152	1	US-08-646-715-17	Sequence 17, Appl
15	231.5	12.0	2085	2	US-08-283-917-8	Sequence 8, Appl
16	221.5	12.0	2085	2	US-08-961-716-8	Sequence 8, Appl
17	200	10.9	2272	4	US-09-108-857-1	Sequence 1, Appl
18	195.5	10.6	1211	4	US-09-063-743-2	Sequence 2, Appl
19	195.5	10.6	1211	4	US-09-590-540-2	Sequence 2, Appl
20	192	10.4	2369	4	US-09-302-769-20	Sequence 20, Appl
21	191	10.4	1115	1	US-08-190-802A-19	Sequence 19, Appl
22	191	10.4	1115	1	US-08-473-346-19	Sequence 19, Appl
23	191	10.4	1115	4	US-08-473-346-19	Sequence 19, Appl
24	191	10.4	1115	4	US-08-473-346-19	Sequence 19, Appl
25	187	10.2	2186	4	US-08-487-072A-19	Sequence 19, Appl
26	187	10.2	2558	4	US-09-184-001-1	Sequence 19, Appl
27	187	10.2	7881	2	US-08-751-189-1	Sequence 1, Appl
28	187	10.2	7881	2	US-09-060-836-1	Sequence 1, Appl
29	187	10.2	7881	4	US-09-184-445-1	Sequence 1, Appl
30	185.5	10.1	1766	5	PCT-US93-00601-1	Sequence 1, Appl
31	185.5	10.1	1766	5	PCT-US94-07107A-1	Sequence 1, Appl
32	185.5	10.1	1767	4	US-08-083-945C-1	Sequence 1, Appl
33	185	10.1	1246	4	US-09-302-769-22	Sequence 2, Appl
34	185	10.1	7886	2	US-08-751-189-2	Sequence 2, Appl
35	185	10.1	7886	2	US-09-060-836-2	Sequence 2, Appl
36	185	10.1	7886	2	US-09-184-445-2	Sequence 2, Appl
37	182.5	9.9	1517	4	US-09-180-783-1	Sequence 1, Appl
38	182	9.9	1394	4	US-09-147-826B-1	Sequence 1, Appl
39	182	9.9	2119	3	US-09-032-372-6	Sequence 6, Appl
40	179.5	9.8	1368	3	US-08-707-339E-1	Sequence 1, Appl
41	179	9.7	2359	1	US-08-188-582-4	Sequence 4, Appl
42	179	9.7	2359	1	US-08-646-715-4	Sequence 4, Appl
43	179	9.7	2481	3	US-08-899-578-1	Sequence 1, Appl
44	175.5	9.5	1020	3	US-08-707-339E-3	Sequence 1, Appl
45	174	9.5	2369	2	US-08-883-534-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-828-922-2
Sequence 2, Application US/08828922
Patent No. 5834240

GENERAL INFORMATION:
APPLICANT: Olga, Bandman
TITLE OF INVENTION: TRANSFORMING GROWTH FACTOR-B RECEPTOR
TITLE OF INVENTION: ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,922
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0258 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166


```

; TELEEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1820 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT01
; CLONE: 753423
; US-08-828-922-2

Alignment Scores:
Pred. No.: 1,21e-209 Length: 1820
Score: 1779.50 Matches: 337
Percent Similarity: 97.15% Conservative: 4
Best Local Similarity: 96.01% Mismatches: 9
Query Match: 96.71% Indels: 1
DB: 2 Gaps: 1

US-09-856-836-2 (1-351) x US-08-828-922-2 (1-1820)

QY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp** 20
DB 259 ATGGAATGAGACAGACGGCGCTCACCTGTTGACACACGGACCGCGTGTGATTTG 318
QY 21 AlaPheSerGlyIleThrProTyGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
DB 319 GCCCTTCAGTGCATCAGCCCTTATGGGTATTTCTTAATCAGCGCTTCAAGATGTAA 378
QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
DB 379 CCTATGCTACCCCAAGGAGATACAGAGACCTGATGGAACATTTTGGGTCAATAAGT 438
QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAsp 80
DB 439 GCTGTTGGCGTGCACACTGAATTAAGATGCCACCAAGCAGCTACACGCTGCAGAT 498
QY 81 PheThrAlaLysValThrAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
DB 499 TTCACAGCCAAAGTGGAGATGCTCTTCAGAGATGATGATGACCCCTGGCTCATTA 558
QY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrLeu**ThrGlyGlyGln 120
DB 559 CACATTTGCAAGACTGCTGATTTCCAGCAGATGATTAATTGTTAAACGGGGGACAG 618
QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluIleSer 140
DB 619 GATTAACCTTACGCAATATGACTTGAACAACTGAAGCAGAACTTAAGATTAGT 678
QY 141 GlyHisThrSerGlyIleLysLysAlaLeuTyrCysSerAspAspLysGlnIleLeuSer 160
DB 679 GGTCTACTCTCGGTATTAATAAAGCTCTGTGTGCTGACATGAGAAACAGATTTCT 738
QY 161 AlaAspAspLysThrValArgLeuThrAspHisAlaThrMetThrGluValLysSerLeu 180
DB 739 GCTGATGACAAAGCTGCTGACTTGGGATCATGCTACTATACAGAAAGTGAATCTCT 798
QY 181 AsnPheAsnMetSerValSerSerMetGluTyrIleProGluGlyGluIleLeuValIle 200
DB 799 AATTATTAATGCTGTTAGTATGATGCAATATATTCCTCGACGAGAGATTTGGTTTA 858
QY 201 ThrTyrGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPhe 220
DB 859 ACTTATGAGAGATCTATCTCTTTTCTTATGTCAGTAAGTTAGCCCAATTAATCTTT 918
QY 221 GluAlaProAlaThrIleAsnSerAlaSer***HisProGluLysGluPheLeuValAla 240
DB 919 GAAGCTCTGCAACCAATCAATCTGCATCTCTCACTGAGAAAGAAATTTCTGTGGA 978
QY 241 GlyGlyGluAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSer 260
DB 979 GCGGCTGAAGATTTTAACCTTTATTAAGTATGATTAATTAAGTGAACAAATTAAGATCC 1038

```

```

QY 261 TyrLysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeuTyr 280
DB 1039 TACAAGGAGACACTTGGTCTCTATTCACGTGTGATTTACTCTGATGAGAACTCTAT 1098
QY 281 AlaSerGlySerGluAspGlyThrLeuArgLeuThrArgGlnThrValAlaGlyLysThrTyr 300
DB 1099 GCCAGTGGTTCAGAGATGAAACATTTGAGACTATGCGCAACTGTGTAGCAAAACCTAT 1158
QY 301 GlyLeuThrPlyCysVal***ProGluGluAspSerGlyGluLeuAlaLysProLysIle 320
DB 1159 GGCCCTTTGGAAATGTGTCTTCTGACAGAGATTAAGTGTGATGCGCAACCAAGATT 1218
QY 321 GlyPheProGluThrAlaGluGluGluLeuAlaGluGluIleAlaSerGluAsnSerAsp 340
DB 1219 GGTTTTCCAGACAAACAGAAAGAGAGCTA---GAAGAATGTCTTACAGAAATTCAGAT 1275
QY 341 SerIleTyrSerSerThrProGluValLysAla 351
DB 1276 TGCATCTTTCCTTCAGCTCTGATGTTAAGGCC 1308

```

RESULT 2

```

US-09-040-984-36
; Sequence 36, Application US/09040984
; Patent No. 6210863
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEEX:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 686 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-040-984-36

```

```

Alignment Scores:
Pred. No.: 4.04e-43 Length: 686
Score: 425.00 Matches: 107
Percent Similarity: 83.97% Conservative: 3
Best Local Similarity: 81.68% Mismatches: 21
Query Match: 23.10% Indels: 7
DB: 4 Gaps: 0

US-09-856-836-2 (1-351) x US-09-040-984-36 (1-686)

```

```
OY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
DB 277 ATGGCAATNAGACAGACCGCGCTGCTCTGGGACACCGCCGCGTGGTATTG 336
OY 21 AlaPheSerGlyIleThrProTyrgIyThrPheLeuIleSerAlaCysLysAspGlyLys 40
DB 337 GCCTTCAGTGGCATCAC-CCTTATGGGTATTCTTATCATCAGCGCTTGCMAAGATGGTTAA 395
OY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
DB 396 CCTATGCTACGCCAGGAGATACAGACAGACTGATGGAACATTTTGGGGCTCAAGCT 455
OY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaIleThrAlaAlaAsp 80
DB 456 CT-GTTTGGGGTGCACACTGATATGATGATGACCAAGACGCTACACAGCTTCAGAT 514
OY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
DB 515 TTCACAGCCCAAGTGGATGCTCTCAGAGANATNATGATATACCTG-CCTCATTA- 572
OY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrlleu***ThrGlyGlyGln 120
DB 573 CACATTTGTCAGAAATGTGATTTCCC-CAGCATAT-TATATTTGTTTACCGGGGAGANAG 630
OY 121 AspLysLeuLeuArgIleTyrrAspLeuAsnLys 131
DB 631 GATTA-CTGTTTNCNMTATTTAATTGAACAA 662

RESULT 3
US-09-123-912-36
: Sequence 36, Application US/09123912A
: Patent No. 6312695
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
: FILE REFERENCE: 210121.455C1
: CURRENT APPLICATION NUMBER: US/09/123,912A
: CURRENT FILING DATE: 1998-07-27
: PRIOR APPLICATION NUMBER: 09/040,802
: PRIOR FILING DATE: 1998-03-18
: NUMBER OF SEQ ID NOS: 114
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 36
: LENGTH: 686
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (222)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (224)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (237)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (264)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (285)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (548)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (551)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (628)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
```

```
: LOCATION: (643)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (645)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (665)
: OTHER INFORMATION: where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (674)
: OTHER INFORMATION: where n is a, c, g or t
US-09-123-912-36

Alignment Scores:
Pred. No.: 4,04e-43 Length: 686
Score: 425.00 Matches: 107
Percent Similarity: 83.97% Conservative: 3
Best Local Similarity: 81.68% Mismatches: 21
Query Match: 23.10% Indels: 7
DB: 4 Gaps: 0

US-09-856-836-2 (1-351) x US-09-123-912-36 (1-686)

OY 1 MetAlaMetArgGlnThrProLeuThrCysSerGlyHisThrArgProValAlaAsp*** 20
DB 277 ATGGCAATNAGACAGACCGCGCTGCTCTGGGACACCGCCGCGTGGTATTG 336
OY 21 AlaPheSerGlyIleThrProTyrgIyThrPheLeuIleSerAlaCysLysAspGlyLys 40
DB 337 GCCTTCAGTGGCATCAC-CCTTATGGGTATTCTTATCATCAGCGCTTGCMAAGATGGTTAA 395
OY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGly 60
DB 396 CCTATGCTACGCCAGGAGATACAGAGACTGATGGAACATTTTGGGGCTCAAGCT 455
OY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaIleThrAlaAlaAsp 80
DB 456 CT-GTTTGGGGTGCACACTGATATGATGATGACCAAGACGCTACACAGCTTCAGAT 514
OY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
DB 515 TTCACAGCCCAAGTGGATGCTCTCAGAGANATNATGATATACCTG-CCTCATTA- 572
OY 101 HisIleValLysThrValAspPheThrGlnAspSerAsnTyrlleu***ThrGlyGlyGln 120
DB 573 CACATTTGTCAGAAATGTGATTTCCC-CAGCATAT-TATATTTGTTTACCGGGGAGANAG 630
OY 121 AspLysLeuLeuArgIleTyrrAspLeuAsnLys 131
DB 631 GATTA-CTGTTTNCNMTATTTAATTGAACAA 662

RESULT 4
US-09-643-597-36
: Sequence 36, Application US/09643597
: Patent No. 6426072
: GENERAL INFORMATION:
: APPLICANT: Wang, Tonglong
: APPLICANT: Fan, Liqun
: APPLICANT: Kalos, Michael D.
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Hosken, Nancy
: APPLICANT: Fanger, Gary R.
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Henderson, Robert A.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.455C1
: CURRENT APPLICATION NUMBER: US/09/643,597
: CURRENT FILING DATE: 2000-08-21
: NUMBER OF SEQ ID NOS: 369
```

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 36
LENGTH: 686
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(686)
OTHER INFORMATION: n = A,T,C or G
US-09-643-597-36

Alignment Scores:
Pred. No.: 4.04e-43 Length: 686
Score: 425.00 Matches: 107
Percent Similarity: 83.97% Conservative: 3
Best Local Similarity: 81.68% Mismatches: 21
Query Match: 23.10% Indels: 7
DB: 4 Gaps: 0

US-09-856-836-2 (1-351) x US-09-643-597-36 (1-686)

QY 1 MetAlaMetArgInThrProLeuThrCysSerGlyHisThrArgProValIValAsp*** 20
DB 277 ATGGCAATNAGACAGACGCCGCTCCTGCTGCGACACGCCGCCGCTGATTG 336
QY 21 AlaPheSerGlyIleThrProTyrGlyTyrPheLeuIleSerAlaCysLysAspGlyLys 40
DB 337 GCCTTCAGTGCATCAC-CCTTATGGGTATTTCTTATTCAGCGCTGCCAAGATGTTAA 395
QY 41 PrometLeuArgGlnGlyAspThrGlyAspTrpIleGlyThrPheLeuGlnHisGly 60
DB 396 CCTATGTTAGCGACGAGATACAGGAGCTGATGGAACATTTTGGGCTCTTAAAGT 455
QY 61 AlaValTrpGlyAlaThrLeuAsnLysAspAlaThrLysAlaIleThrAlaIleAsp 80
DB 456 CT-GTTGGGGTGCACACGATTAAGATGCCACCAAGCAGTACAGCAGTGCAGAT 514
QY 81 PheThrAlaLysValTrpAspAlaValSerGlyAspGluLeuMetThrLeuAlaHisLys 100
DB 515 TTTCACAGCCCAAGTGGTGGATGCTGCTCAGAGATTAATGATTAACCTG-GCTCATTA- 572
QY 101 HisIleValLysThrValAspPheThrGlnAspSerSerGlnLeu***ThrGlyGln 120
DB 573 CACATGTCAGAGATGGAATTCCTCC-CAGGATAT-TATTATTTGTTTACCGGGGAGAG 630
QY 121 AspLysLeuLeuArgIleTyrAspLeuAsnLys 131
DB 631 GATTA-CTGTTTCCNCTATTATTATTAATGACAA 662

RESULT 5
US-09-690-364-10
Sequence 10, Application US/09690364
Patent No. 6468795
GENERAL INFORMATION:
APPLICANT: Hong Zhang
APPLICANT: Andrew T. Walt
TITLE OF INVENTION: ANTISENSE MODULATION OF APAF-1 EXPRESSION
FILE REFERENCE: RTS-0190
CURRENT APPLICATION NUMBER: US/09/690,364
NUMBER OF SEQ ID NOS: 100
SEQ ID NO 10
LENGTH: 5152
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (386)...(4302)
US-09-690-364-10

Alignment Scores:
Pred. No.: 1.08e-23 Length: 5152
Score: 279.00 Matches: 83

Percent Similarity: 44.22% Conservative: 51
Best Local Similarity: 27.39% Mismatches: 143
Query Match: 15.16% Indels: 26
DB: 4 Gaps: 8

US-09-856-836-2 (1-351) x US-09-690-364-10 (1-5152)

QY 13 HisThrArgProValIValAsp***AlaPheSerGlyIleThrProTyrGlyTyrPheLeu 32
DB 2392 CACACAGATGCTGTTTACCAAGCGCTGTTTCT-----CAGGATGGCCAGAGATA 2442
QY 33 IleSerAlaCysLysAspGlyLysPrometLeuArgGlnGlyAspTrpIle 52
DB 2443 GCTTCCTTGGGGCTGATTAACCTTACAGGTGTTCAAAAGCCGAGACGAGAACTT 2502
QY 53 GlyThrPheLeuGlnHisLysGlyAlaValTrpIleThrLeuAsnLysAspAlaThr 72
DB 2503 CTTCACATTAAAGCTCATGAAGATGAGGTGCTGCTGCGGCTTCTCTCAGACAGAGT 2562
QY 73 LysAlaIleThrAlaIleAlaAspPheThrAlaLysValTrpAspAlaValSerGlyAsp 92
DB 2563 TACATAGCAGACCTGCTCAGCGGATTAAGAGTTTGGATTCTCGCAGCGGAG 2622
QY 93 GluLeuMetThrLeu---AlaHisLysHisIleValLysThrValAspPheThrGlnAsp 111
DB 2623 CTTCGACACACCTACGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 2682
QY 112 SerAsnTyr-----Leu***ThrGlyGlnAspLysLeuArgIleTyrAspLeu 129
DB 2683 AGTAAACACCTTCTCTTGGCCACCGGCAAAATATTTCTTCCAAAGCTTGGGATTG 2742
QY 130 AsnLysProGlnAlaGluProLysGluIleSerGlyHisThrSerGlyIleLysLysAla 149
DB 2743 AATCAAAAGAAATGCA---AATACCATGTTTGTGCACAGACAGCAGCAGCAGTGC 2799
QY 150 LeuTrpCysSerAspAspLysGlnIleLeuSerAlaAsp---AspLysThrValArgLeu 168
DB 2800 AGGTTCTCACACGACGATGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2859
QY 169 TrpAspHisAlaThrMetThrGlnValLysSerLeuAsnPheAsnMetSerValSerSer 188
DB 2860 TGGATGTGAGATACAGCAAGACAGAGAAAGCATTAATGGAAGCGCTTCTCTGAGT 2919
QY 189 MetGluTyrIleProGlu----- 194
DB 2920 TCACAAAGCCCTCCAGAGATGAGGTGATGCTGAAGTGTCTGCTGCTGCTGAGAT 2979
QY 195 ---GlyGluIleLeuValIleThrTyrGlyArgSerIleAlaPhe-----HisSerAla 211
DB 2980 GGTGACAAATTAATATGAGGACGAGCAAAAAACAAAGTCTCTTTTGAATTAATCACTAGT 3039
QY 212 ValSerLeuGluProIleLysSerPheGluAlaProAlaThrIleAsnSerAlaSer*** 231
DB 3040 GGCCATTTGGCAGAGATCCACAGGCCAT---CACAGCACATCCAGTACTGACTTC 3096
QY 232 HisProGluLysGluPheLeuValAlaGlyGlyLysAspPheLysLeuTyrLysAsp 251
DB 3097 TCCCTATGACCATTTGGCTGATGATGCCCTGCTCAGTCTGCTGCTGCTGCTGCTGCTG 3156
QY 252 TyrAsnSerGlyGluGluLeuGlnSerTyrLysGlnHisPheGlyProIleHisCysVal 271
DB 3157 ATAAGCTCCGCTTAAAGGTGGCGAGCTGCAGAGACATTGAGTTGGTTACAGGGTGTG 3216
QY 272 ArgPheSerProAspGlyGluLeuTyrAlaSerGlySerGluAspGlyThrLeuArgLeu 291
DB 3217 ATGTTTCTCCGATGGCTCTCATTTTTCACACTCTGTATGACCAACAAATTAAGGCTC 3276
QY 292 TrpGlnThr 294
DB 3277 TGGGAGACA 3285

RESULT 6
US-09-690-364-17

Sequence 17, Application us/09690364

Patent No. 6468795
 GENERAL INFORMATION:
 APPLICANT: Hong Zhang
 APPLICANT: Andrew T. Walt
 TITLE OF INVENTION: ANTISENSE MODULATION OF APAF-1 EXPRESSION
 FILE REFERENCE: RTS-0190
 CURRENT APPLICATION NUMBER: US/09/690,364
 CURRENT FILING DATE: 2000-10-17
 NUMBER OF SEQ ID NOS: 100
 SEQ ID NO 17
 LENGTH: 3747
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(3747)
 US-09-690-364-17

Alignment Scores:

Pred. No.:	2,686-23	Length:	3747
Score:	274.00	Matches:	82
Percent Similarity:	44.19%	Conservative:	55
Best Local Similarity:	26.45%	Mismatches:	146
Query Match:	14.89%	Indels:	27
DB:	4	Gaps:	8

US-09-856-836-2 (1-351) x US-09-690-364-17 (1-3747)

QY 7 ProleuthrCys-SerGlyHisThrArgProValAlaAsp**AlaPheSerGlyIleTh 26
 DB 1821 CCGGTGATGTCGCGCCGCCACACAGTGTGTTACCAAGCCTGCTTTCT----- 1872
 QY 26 rProTyGlyTyrrPheLeuIleSerAlaCysLysAspGlyLysPromeLeuArgIngl 46
 DB 1873 -GAGGATGGTCAGAGATACCTTCTTGTGGAGCTGATAAAAACCTTACAGGCTTCAAGC 1931
 QY 46 yAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaValTrpGlyAlaTh 66
 DB 1932 TGAACGAGGAGAACTTCTAGAAATCAAGGCTCATGAGATGAGAGTCTTGTGTGTC 1991
 QY 66 rLeuAsnLysAspAlaThrIleAlaAlaAlaAlaAspPheThrAlaLysValTr 86
 DB 1992 ATTCTCTACAGATGACAGATTTATACCACTCTCAGTGATGATAAAAAGTGAATTTG 2051
 QY 86 pAspAlaValSerGlyAspGluLeuMetPheLeu---AlaHisLysHisIleValLysTh 105
 DB 2052 GAATTCATGACTGGGGAAGTACTAGACACTGTGAGCACTCAGACCAAGTCAATTG 2111
 QY 105 rValAspPheThrGlnAspSerAsnTrp-----Leu**ThGlyGlyGlnAspLysIle 123
 DB 2112 CAGCCATTTCACCAACAGTATCATCTTCCTTAGCCACTGCTCAAGTCAAGTGCCTT 2171
 QY 123 uLeuArgIleTyrrAspLeuAsnLysProGluAlaGluProLysGluIleSerGlyHisTh 143
 DB 2172 CCTCAAACTTTGGATTTGAATCAAAAAGAAAGTGA--AATACATGTTTGGCATATC 2228
 QY 143 rSerGlyIleLysLysAlaLeuTrpCysSerAspLysGluIleLeuSerAlaAsp-- 162
 DB 2229 AATTCAGTCATCACTGAGATTTTCACCAAGTATGATTAAGCTTTGGCTAGTTGTCACC 2288
 QY 163 -AspLysThrValArgLeuTrpAspHisAlaThrMetThrGluValLysSerLeuAsnPh 182
 DB 2289 TGAATGAACTTAAAGCTTTGGATGCGACATCAAGCAATGAGAGAAAGACATTAAGT 2348
 QY 182 eAsnMetSerValSerMetGluTyrrIleProGlu----- 194
 DB 2349 GAAACAGTTCTTCTTAATTTGGAGACCCCTCAAGAGATATGAGAGTATGAGTGAAGTG 2408
 QY 195 -----GlyGluIleLeuValIleThrTyrrGlyArgSerIleAl 207
 DB 2409 TTGTTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2468

QY 207 aPhe-----HisSerAlaValSerLeuGluProIleLysSerPheGluAlaProAlaTh 225
 DB 2469 TTTTGACATTCATACATGTCGGCTATTTGGAGAAATCCACACCGGCCAT---CACAGCAC 2525
 QY 225 rIleAsnSerAlaSer**HisProGluLysGluPheLeuAlaGlyGlyGlnAspPh 245
 DB 2526 CATTCCAGTACTGATCTCTCCCAACAAACATTTGGCAGTGTGCTTTGCCAGTA 2585
 QY 245 eLysLeuTyrrLysTyrrAspPyrAsnSerGlyGluGluLeuGluSerTyrrLysGlyHisPh 265
 DB 2586 CTGTGTGAGATTTGGATATACAGATCAGCTTCAAGGTGGCTGATTTGAGAGGACATTT 2645
 QY 265 eGlyProIleHisCysValArgPheSerProAspGlyGluLeuTyrrAlaSerGlySerI 285
 DB 2646 AAGTTGGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2705
 QY 285 uAspGlyThrLeuArgLeuTrpGlnThr 294
 DB 2706 TGACACAGACATCATGAGCTCTGGAGACA 2733

RESULT 7

US-09-092-508-1
 Sequence 1, Application us/09092508
 Patent No. 6291643

GENERAL INFORMATION:
 APPLICANT: Henzel, William J.
 TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
 STREET: 3100 No. 6291643 West Center, 90 South Seventh St
 CITY: Minneapolis

STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/092,508
 FILING DATE: 05-JUN-1998
 CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/048,807
 FILING DATE: 05-JUN-1997
 APPLICATION NUMBER: 60/055,258
 FILING DATE: 07-AUG-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Ketelberger, Ph.D., Denise M
 REGISTRATION NUMBER: 33,924
 REFERENCE/DOCKET NUMBER: 11669,6USU1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 TELEX:

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7042 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 578..4159
 OTHER INFORMATION:
 US-09-092-508-1

Alignment Scores:
 Pred. No.: 3,516-22 Length: 7042
 Score: 268.50 Matches: 77

Percent Similarity: 44.03% Conservative: 52
 Best Local Similarity: 26.28% Mismatches: 128
 Query Match: 14.59% Indels: 36
 DB: 4 Gaps: 7

US-09-856-836-2 (1-351) x US-09-092-508-1 (1-7042)

```

OY 7 ProLeuThrCys-SerGlyHisThrArgProValValAsp***AlaPheSerGlyIleTh 26
DB 2365 CCGCTTAGTGTGCGCCGCCACACAGATGCTTTACCATGCTGCTTTCT----- 2416
OY 26 rProTyGlyTyrrPheLeuIleSerAlaCysLysAspGlyLysPrometLeuArgIncl 46
DB 2417 -GAGGATGGTCAGGAATAGCTTTGTGGAGCTGATTAACCTTACAGGTGTCAAAGC 2475
OY 46 yAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaValTrpGlyAlaTh 66
DB 2476 TGAACACAGGAGAACTTCTAGAAATCAAGCTCAAGAGATGAAGTCTTTGTGTGC 2535
OY 66 rLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTr 86
DB 2536 ATTCTCTACAGATACAGATTATAGCAACCTGCTGAGTGATTAATAAGTGAAGATTG 2595
OY 86 pAspAlaValSerGlyAspGlyLeuMetThrLeu---AlaHisLysHisIleValLysTh 105
DB 2596 GAATTCATGACCTGGGAGAACTAGACACACCTATGATGAGCAGCTCAGCAAGTCAATTG 2655
OY 105 rValAspPheThrGlnAspSerAsnTyrr-----Leu***ThrGlyGlyLysAspLys 123
DB 2656 CTGCCATTTTCACCACTAGATCATCTCTCTTACCGCTGCTGCTGCTGCTGCTT 2715
OY 123 uLeuArgTleLysAspLysLeuAsnLysProGluAlaGluProLysGluIleSerGlyHisTh 143
DB 2716 CCTCAAACTTTGGGATTTGATCAAAAAGAAATGTGA---AATACCAGTTTGGTCATAC 2772
OY 143 rSerGlyIleLysLysAlaLeuTyrrPcysSerAspLysGlnIleLeuSerAlaAsp-- 162
DB 2773 AAATTCAGTCAATCAGCTGACAGATTTTCCACAGATGATTAAGCTTTGGCTAGTTCAGC 2832
OY 163 -AspLysThrValArgLeuThrAspHisAlaThrMetThrGlyValLysSerLeuAspPh 182
DB 2833 TGATGGAACTTAAAGCTTTGGGATGCGACATCAGCAAAATGAGAGAAAGCATTAATGT 2892
OY 182 eAsnMetSerValSerSerMetGlyTrpIleProGluGlyIleLeuValIleThrTy 202
DB 2893 GAACAGTTCCTCCATAATTGGAG-----GACCTCAAGAGAGATATGGAAGT 2940
OY 202 rGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProIleLysSerPheGln 222
DB 2918 -----GACCTCAAGAGAGATATGGAAGT 2940
OY 222 aProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAlaGlyL 242
DB 2941 GATAGTGAAGTGTGTGCTGCTGCTGATGATGTCGAACGATTAATGTCGACGCAAAAA 3000
OY 242 yGluAspPheLysLeuTyrrLysTyrrAspTyrrAsnSerGlyGluGlyLeuGlnSerTy 262
DB 3001 T-----AAATCTTTTGTGAAATACAGATCAGCTCAAGCTGATGATGCGAG 3051
OY 262 sGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyLysLeuTyrrAla 282
DB 3052 AGGACATTTTAAGTGGTTCATGCTGTGATGTTTCTCTGATGATCATCATTTTGGAC 3111
OY 282 rGlySerGlyAspGlyThrLeuArgLeuTrpGlnThr 294
DB 3112 ATCTTCTGATGACAGCAATCAGCTCTGGAGAGACA 3148

```

RESULT 8
 US-09-435-115-1
 : Sequence 1, Application US/09435115
 : Patent No. 6346607
 : GENERAL INFORMATION:
 : APPLICANT: Henzel, William J.

TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
 STREET: 3100 No. 634607 West Center, 90 South Seventh St
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/435,115

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/092,508

FILING DATE:

APPLICATION NUMBER: 60/055,258

FILING DATE: 07-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kettelberger, Ph.D., Denise M

REGISTRATION NUMBER: 33,924

REFERENCE/DOCKET NUMBER: 11669,605U1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7042 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 578...4159

OTHER INFORMATION:

US-09-435-115-1

US-09-856-836-2 (1-351) x US-09-435-115-1 (1-7042)

Alignment Scores:
 Pred. No.: 3,51e-22 Length: 7042
 Score: 268.50 Matches: 77
 Percent Similarity: 44.03% Conservative: 52
 Best Local Similarity: 26.28% Mismatches: 128
 Query Match: 14.59% Indels: 36
 DB: 4 Gaps: 7

```

OY 7 ProLeuThrCys-SerGlyHisThrArgProValValAsp***AlaPheSerGlyIleTh 26
DB 2365 CCGCTTAGTGTGCGCCGCCACACAGATGCTTTACCATGCTGCTTTCT----- 2416
OY 26 rProTyGlyTyrrPheLeuIleSerAlaCysLysAspGlyLysPrometLeuArgIncl 46
DB 2417 -GAGGATGGTCAGGAATAGCTTTGTGGAGCTGATTAACCTTACAGGTGTCAAAGC 2475
OY 46 yAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaValTrpGlyAlaTh 66
DB 2476 TGAACACAGGAGAACTTCTAGAAATCAAGCTCAAGAGATGAAGTCTTTGTGTGC 2535
OY 66 rLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTr 86
DB 2536 ATTCTCTACAGATACAGATTATAGCAACCTGCTGAGTGATTAATAAGTGAAGATTG 2595
OY 86 pAspAlaValSerGlyAspGlyLeuMetThrLeu---AlaHisLysHisIleValLysTh 105
DB 2596 GAATTCATGACCTGGGAGAACTAGTACACACCTATGATGACCACTGAGACAGATCAATTG 2655

```

OY	105	rValAsp	hPhc	hGln	Ser	Asn	Trp	-----	Leu	***	Thr	Gly	Gln	Asp	Leu	123							
Db	2656	CTGC	CAAT	TCAC	CCAA	CAG	AGT	GCAT	CTCT	CTT	CTT	AGC	CACT	GGGT	CAAG	CTG	TT						
OY	123	uLeu	Ar	Gle	Tr	Asp	Leu	Asn	Trp	Pro	Glu	Ala	Glu	Pro	Gly	Ser	Gln	His	Thr	143			
Db	2716	CCT	CAAA	CTT	GGG	ATT	TGA	AT	CAAAA	AGAA	TA	TGCA	---	AA	TAC	CAAT	TTT	GGC	CAAT	TC	2717		
OY	143	rSer	Gly	Leu	Leu	Sys	Ala	Leu	Trp	Cys	Ser	Asp	Asp	Leu	Gln	Leu	Ser	Ala	Asp	--	162		
Db	2773	AAA	TTCA	GTCA	TCA	TAC	GCAG	ATTTT	CAC	CA	GCAG	ATGA	TAAG	CTTT	GGC	TGA	TTG	TTT	CA	GC	2832		
OY	163	-Asp	Tr	hVal	Ar	Gle	Tr	Asp	Leu	Trp	Asp	His	Ala	Thr	Met	Thr	Glu	Val	Ser	Leu	Asn	182	
Db	2833	TGA	TGCA	CAAC	CTT	TAA	AGCT	TTT	GGG	AT	TGC	GCAG	CAAT	CAG	CAAA	TAG	AGAG	GA	AAAA	GCA	TTAA	AGT	2892
OY	182	eAsn	Met	Ser	Val	Ser	Ser	Met	Glu	Trp	Leu	Pro	Glu	Glu	Glu	Leu	Val	Leu	Thr	Trp	202		
Db	2893	GAA	ACAG	TTCT	TCT	CTA	TAAT	TTG	GAG	-----											2917		
OY	202	rGly	Ar	Ser	Tr	Leu	Ala	hPhe	hHis	Ser	Ala	Val	Ser	Leu	Glu	Pro	Leu	Ser	Phe	Glu	Ala	222	
Db	2918	-----																					
OY	222	aPro	Ala	Thr	Tr	Leu	Ser	Ala	Ser	***	His	Pro	Glu	Sys	Glu	Pro	Met	Leu	Val	Ala	Gly	242	
Db	2941	GAT	TGCT	CAAG	CTT	GTTC	CGG	CTCT	CTG	ATG	GTGC	CAAC	AGAT	TAAT	GC	CGC	CA	AAAA				3000	
OY	242	yGlu	Asp	Phe	Leu	Sys	Leu	Trp	Trp	Asp	Trp	Asn	Ser	Gly	Glu	Glu	Leu	Asn	Ser	Trp	Leu	262	
Db	3001	T-----																					
OY	262	sGly	His	Phe	hCys	Pro	Leu	hHis	Cys	Val	Ar	Gle	Tr	Asp	Ser	Pro	Asp	Ser	Gly	Leu	Trp	Ala	282
Db	3052	AGG	CACT	TTAA	GTG	GGT	TCAT	GCAT	GGT	GTG	ATG	CTT	TTCT	CTC	TGAT	GAT	CA	TAT	TTT	GAC		3111	
OY	282	rGly	Ser	Glu	Asp	Gly	Tr	hLeu	Ar	Gle	Tr	Asn	Ser	Pro	Asp	Ser	Gly	Leu	Trp	Ala	2		

```

: LENGTH: 7042
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (578)..(4162)
: US-09-690-364-21

Alignment Scores:
Pred. No.: 3.54e-22 Length: 7042
Score: 268.50 Matches: 77
Percent Similarity: 44.03% Conservative: 52
Best Local Similarity: 26.28% Mismatches: 128
Query Match: 14.59% Indels: 36
DB: Gaps: 7

US-09-856-836-2 (1-351) x US-09-690-364-21 (1-7042)

OY 7 Proleuthrcys-serglyhsthrargprovalasp**Alapheserglyileth 26
Db 2365 CCGCTTACTGTCGCGCCACACAGATGCTTACCATGCGCTTCT----- 2416
OY 26 rProtyrglytyrphleuileseralacyslyaspglyysprometleuarglngl 46
Db 2417 -GAGGATGTCAGAGATAGCTTCTTGTGGAGCTGATAAACCTTACAGAGTTCAGAAC 2475
OY 46 yAspThrglyasprrpleglythrphleuuglyhstysglyalavaltpgylalath 66
Db 2476 TGAACACAGAGAGAAACCTTCAAGAAATCAAGCGCTCATGAGATGAGAGTGTGTGTC 2535
OY 66 rleuanslyaspalathrlsAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 86
Db 2536 ATTCCTACAGATGACATATTATAGCAACCTGCTCAGTGATTAAGAGTGAAGATTGG 2595
OY 86 pAspAlaValserglyaspgluleumethleu---Alahlslyshsillevallysth 105
Db 2596 GAATTCCTATGACTGGGAGACAGTACACACCTATGATGAGCAGTACAGCAAGTCAATTG 2655
OY 105 rValaspherthrlnaspsersantyr-----Leu**Thrglyglylaspysle 123
Db 2656 CTGCCATTTCACCAACAGTATGATCATCTTCTTACGCACTGGTCAAGGACCTCT 2715
OY 123 uLeuArgiletyrAspLeuanslysproglnuagluProlysgluileserglyhsth 143
Db 2716 CCTCAACTTGGATTGATGATCAAAAGAAATGTGCA--AATACCATTGTCATATC 2772
OY 143 rserglyllelyslsAlaLeuTPCysersaspslysglnlleleuSerAlaasp-- 162
Db 2773 AATTCAGTCAATCACTGAGATTTTCACCAAGATGATTAAGCTTTGGCTAGTGTTCAGC 2832
OY 163 -AspLysThrValArgLeuTPAspHisAlaThrMetThrGluValLysSerLeuansph 182
Db 2833 TGAATGAACCTTAAGCTTGGATCGCACATCAGCAAAATGAGAGAGAAAGCATTAATGA 2892
OY 182 eAsnMetSerValSerSerMetGluTyrIleProgluglyluilleuValIlethrty 202
Db 2893 GAAACAGTCTTCTTAATTTGGAG----- 2917
OY 202 rgllyargserIleAlaPhenIleSerAlaValSerLeuGluProIleLysSerPheglua 222
Db 2918 -----GACCCCAAGAGATGTGAAG 2940
OY 222 aProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAlaGlygl 242
Db 2941 GATAGTGAAGTGTTCGGTCTGCTGATGTCAGAGATGAATGATGTCAGCAACAAAAA 3000
OY 242 ygluAspPheLysLeuTyrLysTyrAspIyrAsnSerlygluGluLeuGlnserlyrly 262
Db 3001 T-----AAATCTTTTGTGGAATACACACTCAGCTTCAAGGTGGCTGATTCAG 3051
OY 262 sGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyluLeuTyrAlaSe 282
Db 3052 AGGACATTTAAGTGGTTCATGATGTGATGTTTCTCCCTGATGATCATCATTTTTCAG 3111

```

```

OY 282 rGlySerGluAspGlyThrLeuArgLeuTPGlnThr 294
Db 3112 ATCTTGATGACCAAGACATCAGGCTCTGGAGACA 3148

RESULT 11
US-09-692-508-15
: Sequence 15, Application US/09092508
: Patent No. 6291643
: GENERAL INFORMATION:
: APPLICANT: Henzel, William J.
: TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
: STREET: 3100 No. 6291643 West Center, 90 South Seventh St
: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/092,508
: FILING DATE: 05-JUN-1998
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 60/048,807
: FILING DATE: 05-JUN-1997
: APPLICATION NUMBER: 60/055,258
: FILING DATE: 07-AUG-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Ketelberger, Ph.D., Denise M
: REGISTRATION NUMBER: 33,924
: REFERENCE/DOCKET NUMBER: 11669,605U1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-332-5300
: TELEFAX: 612-332-9081
: TELEX:
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7075 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 578...4192
: OTHER INFORMATION:
: US-09-092-508-15

Alignment Scores:
Pred. No.: 3.54e-22 Length: 7075
Score: 268.50 Matches: 77
Percent Similarity: 44.03% Conservative: 52
Best Local Similarity: 26.28% Mismatches: 128
Query Match: 14.59% Indels: 36
DB: Gaps: 7

US-09-856-836-2 (1-351) x US-09-092-508-15 (1-7075)

OY 7 Proleuthrcys-serglyhsthrargprovalasp**Alapheserglyileth 26
Db 2398 CCGCTTACTGTCGCGCCACACAGATGCTTACCATGCGCTTCT----- 2449
OY 26 rProtyrglytyrphleuileseralacyslyaspglyysprometleuarglngl 46
Db 2450 -GAGGATGTCAGAGATAGCTTCTTGTGGAGCTGATTAACCTTACAGGTTCAGAAC 2508

```



```

OY 46 yAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaValTrpGlyAlaThr 66
DB 2509 TGAACAGAGAGAGAACTTCTGAATCAAGCCTCAAGTGAAGTCTTGTGTGC 2568
OY 66 rLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTr 86
DB 2569 ATTCTCTACAGATGACAGATTTATAGCACTGCTCAGGGATTAATAAGTAAAGATTG 2628
OY 86 pAspAlaValSerGlyAspGlyLeuMetThrLeu--AlaHisLysHisIleValLysTh 105
DB 2629 GAATTCATGAGCTGGGAGAACTAGTACACACTATGATGAGCAGCAGCAAGTCAATTG 2688
OY 105 rValAspPheThrGlnAspSerAsnTyL--Leu--ThrGlyGlyGlnAspLysLe 123
DB 2689 CTGCCATTTCACCAACAGTATGATCATCTCTCTTACCCACTGGCTCAAGTACTGCTT 2748
OY 123 uLeuArgIleTyAspLeuAsnLysProGlnuAlaGlnProLysGlnIleSerGlyHisTh 143
DB 2749 CCTCAAACTTGGGATTGAAATCAAAAAGAAATGTGA--AATACCATGTTGGTCATAC 2805
OY 143 rSerGlyIleLysLysAlaLeuTrpCysSerAspLysGlnIleLeuSerAlaAsp-- 162
DB 2806 AAATTCAGTCAATCACTGACAGATTTTTCACCAAGATGATGAAGCTTTGGCTAGTTGCAGC 2865
OY 163 -AspLysThrValArgLeuTrpAspHisAlaThrMetThrGlnuValLysSerLeuAspTh 182
DB 2866 TGATGAGACCTTAAAGCTTTGGGATGCCATCCATCCAAATATGAGAGAAACATTAATGT 2925
OY 182 eAsnMetSerValSerSerMetGluTrpIleProGlnuLysIleuValIleThrTy 202
DB 2926 GAACAGATCTTCTCAATTTGGAG-- 2950
OY 202 rGlyArgSerIleAlaPheHisSerAlaValSerLeuGlnProIleLysSerPheGluAl 222
DB 2951 - 2951
OY 222 aProAlaThrIleAsnSerAlaSer--HisProGlnuLysGlnPheLeuValAlaGlyL 242
DB 2974 GATAGTGAAGTGTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3033
OY 242 yGlnAspPheLysLeuTrpLysTyAspTrpAsnSerGlyGlnuLysGlnuLysSerTyL 262
DB 3034 T-----AAATCTTTTGTGAAATACAGACTCAGCTTCAAAAGTGGCTGATGTCAG 3084
OY 262 sGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGlnuLysValLys 282
DB 3085 AGGACATTTAAGTGGGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3144
OY 282 rGlySerGlnuAspGlyThrLeuArgLeuTrpGlnuThr 294
DB 3145 ATCTTCTGATGAGCAGACAAATCAAGCTCTGGAGACA 3181

```

RESULT 12

```

US-09-435-115-15
; Sequence 15, Application US/09435115
; Patent No. 6346607
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6346607 West Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/435,115
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/092,508
; FILING DATE:
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-Aug-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION/DOCKET NUMBER: 33,924
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 578..4192
; OTHER INFORMATION:
; US-09-435-115-15

```

Alignment Scores:

```

Pred. No.: 3,54e-22 Length: 7075
Score: 268.50 Matches: 77
Percent Similarity: 44.03% Conservative: 52
Best Local Similarity: 26.28% Mismatches: 128
Query Match: 14,598 Indels: 36
DB: 4 Gaps: 7

```

US-09-856-836-2 (1-351) x US-09-435-115-15 (1-7075)

```

OY 7 ProLeuThrCys-SerGlyHisThrArgProValIleAsp--AlaPheSerGlyIleThr 26
DB 2398 CCGCTTAGTGTGCTCCGCCCCACACAGATGCTGTTTCCATGCTGCTGCTTCT-- 2449
OY 26 rProTyGlyTyThrPheLeuIleSerAlaCysLysAspGlyLysProMetLeuArgGlnL 46
DB 2450 -GAGCATGGTCAAGAGAAATAGCTTCTGTGAGCTGATTAACCTTACAGCTTCAAGC 2508
OY 46 yAspThrGlyAspTrpIleGlyThrPheLeuGlyHisLysGlyAlaValTrpGlyAlaTh 66
DB 2509 TGAACAGAGAGAGAACTTCTGAATCAAGCCTCATGAGATGAAGTCTTGTGTGC 2568
OY 66 rLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTr 86
DB 2569 ATTCTCTACAGATGACAGATTTATAGCACTGCTCAGGGATTAATAAGTAAAGATTG 2628
OY 86 pAspAlaValSerGlyAspGlyLeuMetThrLeu--AlaHisLysHisIleValLysTh 105
DB 2629 GAATTCATGAGCTGGGAGAACTAGTACACACTATGATGAGCAGCAGCAAGTCAATTG 2688
OY 105 rValAspPheThrGlnAspSerAsnTyL--Leu--ThrGlyGlyGlnAspLysLe 123
DB 2689 CTGCCATTTCACCAACAGTATGATCATCTCTCTTACCCACTGGCTCAAGTACTGCTT 2748
OY 123 uLeuArgIleTyAspLeuAsnLysProGlnuAlaGlnProLysGlnIleSerGlyHisTh 143
DB 2749 CCTCAAACTTGGGATTGAAATCAAAAAGAAATGTGA--AATACCATGTTGGTCATAC 2805
OY 143 rSerGlyIleLysLysAlaLeuTrpCysSerAspLysGlnIleLeuSerAlaAsp-- 162
DB 2806 AAATTCAGTCAATCACTGACAGATTTTTCACCAAGATGATGAAGCTTTGGCTAGTTGCAGC 2865
OY 163 -AspLysThrValArgLeuTrpAspHisAlaThrMetThrGlnuValLysSerLeuAspTh 182

```

Db 2866 TGATGACCTTAAAGCTTTGGATGCGACATGACGAAATGAGAGAAAGCATTAATGT 2925
 QY 182 easnmetserValserMetgluTyrilleProgluylgluileuVallethrtY 202
 Db 2926 GAAGACATTTCTCTAATATTGGAG----- 2950
 QY 202 rglYarSerilleAlahehSserAlaValSerleuGluproileylsSerPhegluAl 222
 Db 2951 -----GACCTTCAGAGATTTGAAAGT 2973
 QY 222 aproAlaThrileasnsrAlaSer***HisprogluylgluPheleuValAlaaglYgl 242
 Db 2974 GATAGTGAAGCTGTCTGCTGCTGATGTCGACATTAATGTCGACACCAAAAA 3033
 QY 242 ygluAsPheleuylsleuTyrlySTyrAsPyrAsnSerGlygluGluleneuGlusertYrly 262
 Db 3034 T-----AAATCTTTTGTGGAATACAGACTCAGTCAAAAGGTGCTGATTCAG 3084
 QY 262 sglYHisPheleuylsleuThrleuHisCysValArgPheSerProAspGlygluLeuTyrAlaSe 282
 Db 3085 AGGACATTTAAAGTGGTTCATGCTGATGCTGATTTCTTCATGATCATCATTTTGGAC 3144
 QY 282 rglYserGlupAspGlyThrleuArgleuTyrPgluThr 294
 Db 3145 ATCTTCTGATGACGACGACATCAGGCTCTCGAGAGACA 3181

RESULT 13

US-08-188-582-17
 : Sequence 17, Application US/08188582
 : Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
 APPLICANT: Comal, Lucio
 APPLICANT: Dynalact, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TFBS AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188, 582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2152 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA

FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2112
 US-08-188-582-17

Alignment Scores:

Pred. No.:	2,666-19	Length:	2152
Score:	238.50	Matches:	86
Percent Similarity:	40.36%	Conservative:	50
Best Local Similarity:	25.52%	Mismatches:	147
Query Match:	12.96%	Indels:	55
DB:	1	Gaps:	11

US-09-856-836-2 (1-351) x US-08-188-582-17 (1-2152)

QY 54 ThrPheleuGlyHislysGlyAlaValITrPglYAlaThrleuAsnLysAspAlaThrlys 73
 Db 1105 ACATTTCTCAATGCTTAAACAGGCTGTCACTGACAGTGTGATGTCATGATCTTCACTG 1164
 QY 74 AlaAlaThrAlaAlaAspPheThrAlaLysValITrPAspAlaValSerGlyAspGlu 93
 Db 1165 ATTGCTGAGGTTTTCAGATTCACACTGTACAGTGTG---TCGGTAAACCCCAAAAG 1221
 QY 94 LeuMetThrleuAlaHisLysHisLleValLysThrValAspPheThrGlnLysSerAsn 113
 Db 1222 CTTCGATGCTCAACAGACATCAGATCTTATGCTTATGACAAAGATCAGAT----- 1275
 QY 114 Tyrleu**ThrGlyGlyGlnAspLysLeuValArgleuTyrAspPheLysProglu 133
 Db 1276 -----GATGCTTAAAGATCATGAT---GAGAAACAGCA 1311
 QY 134 AlaGluProLysGluLeuSerGlyHisThrSerGlyTyrleuLysAlaLeuTyrPysSer 153
 Db 1312 AGTAGTGAAGATTTGTATGTCACAGGCGCTGTACGACCCAGCTTCACTGCG 1371
 QY 154 AspAspLysGlnleuLeu---SerAlaAspAspLysThrValArgleuTyrPAspHisAla 172
 Db 1372 GATGACACTATCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1431
 QY 173 ThrMetThr---GluValLysSerleuAsnPheAsnMetSerValSerSerMetGluTyr 191
 Db 1432 ACATTTACTGTTGGTGGGATTAAGACACACACATCAGTATGCGACACACATTT 1491
 QY 192 IleProgluylgluileuValle---ThrTyrGlyArgSerilleAlaPheHisSer 210
 Db 1492 TCTCCATATGGAATTAATTTGTGTACAGGGGCCATGACCAAGTACTGCGCTGGCT 1551
 QY 211 AlaValSerleuGluproileylsSerPhegluAlaPro---AlaThrleasnsrAla 229
 Db 1552 ACAGACCACTATACGCTTTAAGAAATTTGCGCGCATCTGCTGATGATGATGATG 1611
 QY 230 Ser**HisprogluylsGluPheleuValAlaaglYgluAspPheLysleuTyrlys 249
 Db 1612 AGATTCATCCAAATTCATTAATGTTGCTPACGGGCTCTGACAGACACATGCGGCTC 1671
 QY 250 TyrAspPyrAsnSerGlygluGluleneuGlusertYrlySTyrAsnSerPhegluPheleu 269
 Db 1672 TGGACGCTCTGAATGGAATGTAAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1731
 QY 270 CysValArgPheSerProAspGlygluLeuTyrAlaSerGlyserGlyAspGlyThrleu 289
 Db 1732 TCCTTGACATTTTCTCCCAATGAGATTCCTGCTPACAGAGACAGATGACAGAGT 1791
 QY 290 ArgleuTyrGln-----ThrValValglYshThrTyrGly 301
 Db 1792 CTTCCTTGGATTAATGACATGTTGATGTTGAGAAATTAAGGACACACTGATACA 1851
 QY 301 ----- 301
 Db 1852 GTCTGTTACTTACTGTTAGTAGAGATGTAATTTTGGATCAGTTCATGATGATAT 1911
 QY 302 -----LeuTyrLysCysVal**Pro-----GluLysAspSerGlyGlu 314
 Db 302 ----- 314


```

Patent No. 5849557
GENERAL INFORMATION:
APPLICANT: ADACHI, HIDEKI
APPLICANT: TSUJIMOTO, MASAFUMI
APPLICANT: INOUE, KEIZO
APPLICANT: ARAI, HIROYUKI
TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
TITLE OF INVENTION: AND GENE THEREOF
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER &
ADDRESSEE: NEUSTADT, P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/283,917
FILING DATE: 03-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 209943/1993
FILING DATE: 03-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5849557man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2292-030-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: CDS
LOCATION: 844..2073
US-08-283-917-8
Alignment Scores:
Pred. No.: 3,17e-17 Length: 2085
Score: 221.50 Matches: 72
Percent Similarity: 44.83% Conservative: 58
Best Local Similarity: 24.83% Mismatches: 123
Query Match: 12.04% Indels: 37
DB: 2 Gaps: 9
US-09-856-836-2 (1-351) x US-08-283-917-8 (1-2085)
QY 11 SerGlyHisThrArgProValValasp***AlapheserGlyIleThrProTyrGlyTyr 30
DB 1156 AGTGGTCATAGAGAGTCACAGTCATTTCCAT-----CCTGTGTTCACT 1206
QY 31 PheLeuIleSerAlaCysAspGlyLysProMetLeuArgGlnGlyAspThrGlyasp 50
DB 1207 GTTATGCTCTCTCCTCAAGAGATGCTACAAATTAGGTGTGGATTATGAGACTGGAGAT 1266
QY 51 TTPileGlyThrPheLeuGlyHisLysGlyAlaValTTPGlyAlaThrLeuAsnLysasp 70
DB 1267 TTTGAACGAACTCTTAAGGGCATACAGACTCTGTACAGATATTTCATTCGACCACT 1326

```

```

QY 71 AlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValTTPaspAlaValSer 90
DB 1327 GGCAAGCTTCTGGCTTGCATGTTCTCAGATATGACCATTAAGCTATGGATTTCCAGGC 1386
QY 91 GlyAspGluLeuMetThrLeu---AlaHisLysHisIleValLysThrValAspPheThr 109
DB 1387 TTTGATGATCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1446
QY 110 GlnAspSerAsnTyrLeu***ThrGlyGlnAspLysLysLeuValGlyLeuTyrAspLeu 129
DB 1447 CCCAATGAGATCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1506
QY 130 AsnLysProGlnAlaGluProLysGluLysSerGlyHisThrSerGlyIleLysAla 149
DB 1507 CAAATGCTACTGT---GTGAAGACATTCACAGACAGACAGAAAGGCTAGTTGGTGTG 1563
QY 150 LeuTTPCysSerAspAspLysGlnIleLeuSer---AlaAspAspLysThrValArgLeu 168
DB 1564 CGGCCAATCAAGACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1623
QY 169 TTPAspHisAlaThrMetThrGluValLysSerLeuAsnPheAsnMetSerValSerSer 188
DB 1624 TGGCTGTGTAACAACA---AAGGAATGCAAGGCT-----GAGCTTGAGAGA 1665
QY 189 MetGluTyrTleProGluGluGluLeuValIleThrTyrGlyArgSerIleAlaPhe 208
DB 1666 CATGACCATGTGTGAGA----- 1683
QY 209 HisSerAlaValSerLeuGluProLysSerPheGluAlaProAlaThrIleAsnSer 228
DB 1684 -----TGCATTTCTGGGCTCTGAAAGCTCATAT-----TCTTCATCTCTGAA 1728
QY 229 AlaSer***HisProGluLysGlu-----PheLeuValAlaGlyGly 242
DB 1729 GCAACAGCATGTCAGACTAAAGTGCACAACTGGGCCATTTCTACTGTCCGATCC 1788
QY 243 GlnAspPheLysLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluLysTyrLys 262
DB 1789 AGGCACAAAGCTATCAAGATGTGGATGTCACTACGATGCGATGCTTATGACCTGGTG 1848
QY 263 GlyHisPheGlyProIleHisCysValArgPheSerProAspGlyLysLeuTyrAlaSer 282
DB 1849 GGTCAATGATTAAGTGGATCGTGGAGTGTCTGTCCATTTCTGGGGGAAGTTATTGAGT 1908
QY 283 GlySerGlnAspGlyThrLeuArgLeuTyr 292
DB 1909 TGGCTGATGACAAAGCCTGCGGCTGG 1938

```

Search completed: February 6, 2003, 22:29:38
 Job time : 92 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 6, 2003, 13:11:53 ; Search time 69 Seconds

(without alignments)
2411.233 Million cell updates/sec

Title: US-09-856-836-2

Perfect score: 1840
Sequence: 1 MAMRGTPLTCSGHRPVVDX.....EELASNSDSIYSTPEVKA 351

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 408267 seqs, 237001491 residues

Total number of hits satisfying chosen parameters: 816534

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-O=/cgn2_1/USPFO.spool/US09856836/runat_31012003_144239_22239/app_query.fasta_1.519
-DB=Published.Applications.NA -OFMT=fastap -SUFFIX=rmpb -MINMATCH=0.1
-LOOEXT=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62
-TRANS=numa40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09856836 -CGN_1_1_24 -errunat_31012003_144239_22239
-NCPU=6 -ICPU=3 -NO_XIPYX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1347.5	73.2	1430	10	US-09-925-301-520 Sequence 520, App
2	781.5	42.5	1002	9	US-09-938-842A-510 Sequence 510, App
3	747	40.6	437	10	US-09-960-352-8644 Sequence 8644, App
4	517.5	28.1	313	10	US-09-815-343-740 Sequence 740, App

5	517.5	28.1	313	10	US-09-998-598-1947 Sequence 1947, App
6	487	26.5	311	9	US-09-920-455-140 Sequence 140, App
7	445	24.2	469	10	US-09-770-444-220 Sequence 220, App
8	425	23.1	686	10	US-09-735-705-36 Sequence 36, App
9	425	23.1	686	10	US-09-850-716A-36 Sequence 36, App
10	425	23.1	686	10	US-09-897-778-36 Sequence 36, App
11	349	19.0	361	10	US-09-998-598-2482 Sequence 2482, App
12	272.5	14.8	3025	10	US-09-764-853-29 Sequence 29, App
13	268.5	14.6	7042	10	US-09-876-667-15 Sequence 15, App
14	268.5	14.6	7072	10	US-09-876-667-15 Sequence 15, App
15	233.5	12.7	1522	9	US-10-098-841-122 Sequence 122, App
16	227.5	12.4	1378	9	US-10-098-841-121 Sequence 121, App
17	211.5	11.5	2401	10	US-09-729-674-167 Sequence 167, App
18	205.5	11.2	2151	12	US-10-042-417-1 Sequence 1, App
19	201.5	11.0	463	10	US-09-924-035A-278 Sequence 278, App
20	201.5	11.0	1410	9	US-10-109-579A-1 Sequence 1, App
21	200	10.9	262	10	US-09-923-876-3070 Sequence 3070, App
22	199.5	10.8	2551	9	US-10-098-841-285 Sequence 285, App
23	199.5	10.8	2590	9	US-10-098-841-284 Sequence 284, App
24	199.5	10.8	2710	12	US-10-044-090-646 Sequence 646, App
25	199	10.8	4285	12	US-10-044-090-646 Sequence 646, App
26	198.5	10.8	2037	10	US-09-801-368-313 Sequence 313, App
27	198	10.8	1545	10	US-09-917-800A-1690 Sequence 1690, App
28	197.5	10.7	1045	10	US-09-828-310-7 Sequence 7, App
29	196.5	10.7	1881	9	US-09-213-888-20 Sequence 20, App
30	196.5	10.7	1881	9	US-09-328-877A-20 Sequence 20, App
31	196.5	10.7	2001	9	US-09-213-888-26 Sequence 26, App
32	196.5	10.7	2001	9	US-09-328-877A-26 Sequence 26, App
33	196.5	10.7	2010	9	US-09-213-888-24 Sequence 24, App
34	196.5	10.7	2010	9	US-09-328-877A-24 Sequence 24, App
35	196.5	10.7	3550	9	US-09-328-877A-1 Sequence 1, App
36	196.5	10.7	3550	9	US-09-328-877A-1 Sequence 1, App
37	196.5	10.7	3571	9	US-09-213-888-2 Sequence 2, App
38	196.5	10.7	3571	9	US-09-328-877A-2 Sequence 2, App
39	195.5	10.6	1211	9	US-10-119-932-2 Sequence 2, App
40	194	10.5	1288	10	US-09-770-445-18 Sequence 18, App
41	192	10.4	2369	10	US-09-908-805B-20 Sequence 20, App
42	191	10.4	1089	10	US-09-917-800A-1444 Sequence 1444, App
43	191	10.4	1093	10	US-09-969-708-214 Sequence 214, App
44	191	10.4	1093	10	US-09-880-107-2302 Sequence 2302, App
45	187	10.2	2186	9	US-09-943-689A-1 Sequence 1, App

ALIGNMENTS

RESULT 1
US-09-925-301-520
Sequence 520, Application US/09925301
Patent No. US2002052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
FILE OF INVENTION: Nucleic Acids, Proteins and Antibodies
TITLE REFERENCE: PA106
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: US/09/925,301
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 520
LENGTH: 1430
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (104)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (105)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature

Alignment Scores:			
Pred. No.:	6.4e-82	Length:	437
Score:	747.00	Matches:	141
Percent Similarity:	99.31%	Conservative:	3
Best Local Similarity:	97.24%	Mismatches:	1
Query Match:	40.60%	Indels:	0
DB:	10	Gaps:	0

US-09-856-836-2 (1-351)	x	US-09-960-352-8644 (1-437)
QY	138	GIUILESERGIYHISTHRSERGILYILEYLSYSLALEUTRPGYSSERPASPPLYSGLN 157
DB	3	GAATCATGTGGTCACACCTCGTATATAAAAGGCTTTGTGTGCATGAGGATAAACAG 62
QY	158	ILEUSERALASPASPPLYSFTHVALRGLLEUTRPHSPHISALATMRMETHGLVAL 177
DB	63	ATTCCTTTCAGCGGATGATAAACATGTTGTCCTTGGGATCATCTACTATGACGCGAAGTG 122
QY	178	LYSSERLEAAPHASMEISERVALSERSERMETGLUTYRILIERPROGLUGLYGLILE 197
DB	123	AAATCTCTAAATTTTAAATATATCCGTTAGCGATGGAATACATCTCTGAGGAGGAATT 182
QY	198	LEUVALIETHRTYGLVARGSERILELAPHENHISERIALVALSERLEUGLUPROILE 217
DB	183	TTGGTAATTAACCTTAAGGACGATCTATCTCTTTTATATAGTAGGAGTAAGTTTGACCAATT 242

Alignment Scores:	
Pred. No.:	2.94e-54
Score:	517.50
Percent Similarity:	94.29%
Best Local Similarity:	92.38%
Query Match:	28.12%
DB:	10
US-09-856-836-2 (1-351) x US-09-815-343-740 (1-313)	
	Length: 313
	Matches: 97
	Conservative: 2
	Mismatches: 5
	Indels: 1
	Gaps: 1

QY	246	LysLeuTyrLysTyrAspTyrAsnSerGlyGluLeuGlnSerTyrLysGlyHisPhe	265
Db	1	AAACTTTATAGTATGATTATTAATAGTGAGAAATAATAGATCTCTCAAGGGACACTTT	60
QY	266	GlyProIleHisCysValArgPheSerProaspGlyLeuTyrAlaSerGlySerGlu	285
Db	61	GCTCTATTACTGCTGGTAGATTAGTCTCATGAGAAACTCTATGCGACGGGTCA	120
QY	286	AspGlyThrLeuArgLeuTyrGlnThrValValGlyLysThrTyrGlyLeuTyrLysCys	305
Db	121	GATGGAACAATTGACGACTTGGCCAAACTTGGTACGAAAAACGTATGGCTTTGGAAATGT	180
QY	306	Val***ProGluAspSerGlyGluLeuAlaLysProLysIleGlyPheProGluThr	325
Db	181	GTGCTTCTGAAAGAAAGATAGTGGTAGCTGGCAAGCCAAAGATTGGTTTCCAGAACCA	240
QY	326	AlaGluGluGluLeuAlaGluGlnIleAlaSerGlnSubSerAspSerIleTyrSerSer	345
Db	241	ACAGAAAGAGGAGCTA---GAAGAAATTCTCTCAAGAGATTCAAGATTCATCTTCTTCA	297
QY	346	ThrProGluValLys 350	
Db	298	GCTCCTGATGTTAAG 312	


```
; Sequence 1947, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1947
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1947
```

```
Alignment Scores:
Pred. No.: 2,94e-54 Length: 313
Score: 517.50 Matches: 97
Percent Similarity: 94.29% Conservative: 2
Best Local Similarity: 92.38% Mismatches: 5
Query Match: 26.12% Indels: 1
Gaps: 1
```

US-09-856-836-2 (1-351) x US-09-998-598-1947 (1-313)

```
QY 246 LysLeuTyrLysTyrAspTyrAsnSerGlyGluGluSerTyrLysGlyHisPhe 265
DB 1 AAACCTTAATAGTATATATATAGTGAGAGAAATAGAACTCTACAGAGACACTT 60
QY 266 GlyProIleHisCysValAlaPheSerProAspGlyGluLeuTyrAlaSerGlySer 285
DB 61 GGTCTATTCATCTGTGTGAGATTAGCTCTGATGAGAACTTATGCCAGGCTTCA 120
QY 286 AspGlyThrLeuArgLeuTrpGlnThrValAlaGlyThrTyrGlyLeuTrpLysCys 305
DB 121 GATGGAACATTGAGACTATGCGCAACTGTGTAGAAACGATGCGCTTGAAATGT 180
QY 306 Val***ProGluGluAspSerGlyGluLeuAlaLysProLysIleGlyPheProGluThr 325
DB 181 GTGCTCTCTGAAGAAAGATGTGTGAGCTGGCAAGCCAAAGATTGTTTCCAGAGCA 240
QY 336 AlaGluGluGluLeuAlaGluGluLeuAlaSerGluAsnSerAspSerIleTyrSerSer 345
DB 241 ACAGAAAGAGAGCTA---GAAGAAATTCCTTCAGAGATTCAGATTGCATCTTCTTCA 297
QY 346 ThrProGluValLys 350
DB 298 GCTCCTGATGTTAAG 312
```

```
RESULT 6
US-09-920-455-140
; Sequence 140, Application US/09920455
; Patent No. US20020168647A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF HEAD AND NECK CANCER
; FILE REFERENCE: 210121.540
; CURRENT APPLICATION NUMBER: US/09/920,455
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 275
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc-feature
; LOCATION: 115, 199, 201, 235, 238
; OTHER INFORMATION: n = A,T,C or G
US-09-920-455-140
```

```
Alignment Scores:
Pred. No.: 1.48e-50 Length: 311
Score: 487.00 Matches: 90
Percent Similarity: 95.79% Conservative: 1
Best Local Similarity: 94.74% Mismatches: 4
Query Match: 26.47% Indels: 0
Gaps: 0
```

US-09-856-836-2 (1-351) x US-09-920-455-140 (1-311)

```
QY 212 ValSerLeuGluProIleLysSerPheGluAlaProAlaThrIleAsnSerAlaSer*** 231
DB 23 GTAAGTTGGACCCAAATTAATCTTTGAAGCTCTGCAACCAATCAATTCGATCTCTT 82
QY 232 HisProGluLysGluPheLeuValAlaGlyGlyLysAspPheLysLeuTyrLysTyrAsp 251
DB 83 CATCTGAGAAAGAAATTTCTTGTGACGGCGNAGATTTTAACTTTATAGTATGAT 142
QY 252 TyrAsnSerGlyGluGluLeuGluSerTyrLysGlyHisPheGlyProIleHisCysVal 271
DB 143 TATATAGTGAGCAAGAAATTAAGATCTTACAGAGACACTTGGCGCTTATTCCTGNG 202
QY 272 ArgPheSerProAspGlyGluLeuTyrAlaSerGlySerGluAspGlyThrLeuArgLeu 291
DB 203 AGATTAGTCTGTGAGGAGAACTTATGCGCAGNGNCACAAATGGAACATTGAGACTA 262
QY 292 TrpGlnThrValAlaGlyLysThrTyrGlyLeuTrpLysCysVal 306
DB 263 TGGCAAACTGTGTAGGAAAAACGTATGCGCTTGGAAATGTGTG 307
```

RESULT 7

US-09-770-444-220

Sequence 220, Application US/09770444

Patent No. US2002023280A1

GENERAL INFORMATION:

```
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Olang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 220
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-220
```



```

QY 26 rProtyrGlyTyrPheLeuIleSerAlaCysAspGlyLysPrometLeuArgGlnG1 46
Db 2450 -GAGGATGGTCAAGAAATGCTTCTGTGGAGCTGATAAACCTTACAGAGTTCACAAAGC 2508
QY 46 yAspThrGlyAspThrIleGlyThrPheLeuGlnHisLysGlyAlaValThrPGLAla1th 66
Db 2509 TCAACAGAGAGAAACCTTCTGAAATCAAGAGCTATGAGTGAAGTGTCTTGTTCGC 2568
QY 66 rLeuAsnLysAspAlaThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValThr 86
Db 2569 ATCTCTCAAGTACAGATTTATTAACCACTGCTCAGTGAATAAAGTAAACATTTTG 2628
QY 86 pAspAlaValSerGlyAspGluLeuMetThrLeu---AlaHisLysHisLysValLysTh 105
Db 2629 GAATCTATGACTGGGAGACTAGTACACACCTATGATGAGCATGACAGCAAGTCAATTC 2688
QY 105 rValAspPheThrGlnAspSerAsnTyr-----Leu---ThrGlyGlyGlnAspLysLe 123
Db 2689 CTGCCATTTCAACACAGCTAGTCATCATCTCTTACCCACTGGGTCAAGTCACTGCTT 2748
QY 123 uLeuArgIleTyrAspLeuAsnLysProGluAlaGluProLysGluLysSerGlyHisTh 143
Db 2749 CTTCAAACTTTGGGATTTGAAATCAAAAAGAAATGTCGA---AATACCATGTTGGTCAATC 2805
QY 143 rSerGlyLysLysLysAlaLeuTyrCysSerAspAspLysGlnIleLeuSerAlaAsp-- 162
Db 2806 AAATTCATCATCATCATCATCATTTTCCACAGATGATTAAGCTTTGGCTTGTTCAGC 2865
QY 163 -AspLysThrValArgLeuTyrAspHisAlaThrMetThrGlyValLysSerLeuAsnPh 182
Db 2866 TGAATGGAACCTTAAAGCTTTGGATGCCACATCATGCAAAATGAGAGAAACATTAATGT 2925
QY 182 eAsnMetSerValSerSerMetGluTyrIleProGluGlyGluIleLeuValIleThrTy 202
Db 2926 GAACACGTTCTTCTTAAATTTGGAG----- 2950
QY 202 rGlyArgSerIleAlaPheHisSerAlaValSerLeuGluProLysSerPheGluAl 222
Db 2951 -----GACCCTCAAGAGATATGAGAGT 2973
QY 222 aProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAlaGlyG1 242
Db 2974 GATAGTCAAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3033
QY 242 yGluAspPheLysLeuTyrTyrAspTyrAsnSerGlyGluGluLeuLeuLysTyrLy 262
Db 3034 T-----AAATCTTTTGTGGAATACAGACTCAGCTTCAAGAGTGGCTGATTCAG 3084
QY 262 sGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeuTyrAlaSe 282
Db 3085 AGAGCATTTAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3144
QY 282 rGlySerGluAspGlyThrLeuArgLeuThr 294
Db 3145 ATCTCTGATGACACAGCAATCAAGGCTGTGGAGACA 3181

```

RESULT 15
US-10-098-841-122
Sequence 122, Application US/10098841
Publication No. US2002019769A1
GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Duntui
APPLICANT: Wang, Zhilwei

```

APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US2002019769A1el Nucleic Acids and
FILE REFERENCE: 784CIP2
CURRENT APPLICATION NUMBER: US/10/098,841
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
NUMBER OF SEQ ID NOS: 331
SOFTWARE: PL-FL-genes Version 1.0
SEQ ID NO 122
LENGTH: 1522
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (456)..(1061)
US-10-098-841-122

```

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	1.06e-18	23.50	40.18	23.62	1522	77	54	130	65	12

```

US-09-856-836-2 (1-351) x US-10-098-841-122 (1-1522)
QY 52 rLeuGlyThrPheLeuGlnHisLysGlyAlaValThrPGLAlaThrLeuAsnLysAspAla 71
Db 39 CTGTGATCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 86
QY 72 ThrLysAlaAlaThrAlaAlaAspPheThrAlaLysValThr---AspAlaValSer 90
Db 87 -----ACTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 116
QY 91 gLysAspGluLeuMetThrThrLeuAlaHisLysHisLysValLysThrValAspPheThrGln 110
Db 117 GGTAGTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 176
QY 111 AspSerAsnTyrLeu---ThrGlyGlyGlnAspLysLeuLeuArgIleTyrAspLeuAsn 130
Db 177 GATGGCAGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 236
QY 131 LysProGluAlaGluProLysGluLysSerGlyHisThrSerGlyIleLysLysAlaLeu 150
Db 237 CCC---CAAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 293
QY 151 TrpCysSerAspAspLysGlnIleLeuSerAla---AspAspLysThrValArgLeuThr 169
Db 294 TTTCTCCCGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353
QY 170 AspHisAlaThrMetThrGluValLysSer---LeuAsnPheAsnMetSerValSerSer 188
Db 354 GACAAAGAGAGCGGGAATGTGTCCACCTGATTTGTGAGCATGGCGGCTTGTGCACCTAT 413
QY 189 MetGluTyrIleProGluGlyGluIleLeuValIleThrTyrGlyArgSerIleAlaPhe 208
Db 414 GTGACCTTCCACCCAGGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 458
QY 209 HisSerAlaValSerLeuGluProIleLysSerPheGluAla-----GCTGCCGAGATG 478
Db 459 GACACACAGTGAAGGTGTGGAGTGTGGAGTGTGGAGTGTGGAGTGTGGAGTGTGGAGTGTGGAGT 518
QY 223 --ProAlaThrIleAsnSerAlaSer**HisProGluLysGluPheLeuValAlaGly 241

```

```

Db      519  CACAGTGCAGCAGTGCAGCGGCTCTCTTCCACCCGCGGAAACTGACTGATCACAGCC 578
           |||  ::|||  |||  |||||  ::|||::
QY      242  GlyGluAspPheLeuTyrLysTyrAspTyrAsnSerGlyGluGluLeuGluSerTyr 261
           |||  |||  |||  |||  |||  ::
Db      579  TCCAGTGCATCAACCTGAAATCCTGGACCTGATGAGGCGGCTGCTTACACACTC 638
           |||  |||  |||  |||  |||  ::
QY      262  LysGlyHisPheGlyProIleHisCysValArgPheSerProAspGlyGluLeuTyrAla 281
           |||||  |||||  |||  |||||  |||||  ::|||
Db      639  CACGGGATCAGGAGCACCCACTGTGCTTTCAGAAACGGGGAGATATTTTGT 698
           |||||  |||||  |||  |||||  |||||  ::|||
QY      282  SerGlySerGluAspGlyThrLeuArgLeuTrpGlnThr-----ValValGlyLys 298
           |||||  |||  |||  ::|||  ::|||
Db      699  TCTGAGGCTCTGATGACAGTATGTTGGAAGAGTACTTGATATGTT----- 752
           |||||  |||  |||  ::|||  ::|||
QY      299  ThrTyrGlyLeuTrpLysCysVal***ProGluGluAspSerGlyGluLeuAlaLys--- 317
           |||  |||  |||||  ::|||  |||
Db      753  -----GATCATGAGAGAGTCAAGAAAGTG 776
           |||  |||  |||||  ::|||
QY      318  -----ProLysIleGlyPhePro 323
           |||  |||  |||||  ::|||
Db      777  CCGAGGCCCCAGCCACACTGGCCAGCTCCATGGGGAATCTGCCAGAGATGGACTTCCT 836
           |||  |||  |||  |||  |||  ::
QY      324  ---GluThrAlaGluGluGluLeuAlaGluIleAlaSerGluAsnSerAspSerIle 342
           |||  ::|||  ::|||  ::|||  ::|||
Db      837  GTCCCCCAGCAGAGGAGAGTGTGAGTCTGTGCAGAGCCAGCCAGAGACCCGTG 896
           |||  ::|||  ::|||  ::|||
QY      343  TyrSerSerThrProGlu 348
           |||  |||||  ::|||
Db      897  -----AGTGTGCCCCAG 908
           |||  |||||  ::|||

```

Search completed: February 7, 2003, 00:01:23
 Job time : 86 secs

OY 326 TTCACAGCCAAAGTATGGGATGGCGGTCTCAGAGGATGAATGATGACCTGGCTCTATAG 38
 |||||
 Db 39 TTCACAGCCAAAGTGGGATGCTGTCTCAGAGATGAATTGATGACCTGGCTATAAA 98
 |||||
 OY 386 CACATTGTCAGAGCTGTGGATTTTACACAGGATAGCAATTACCTGTAACCTGGGACAG 445
 |||||

[illegible]

RESULT 8
US-09-920-455-140

Query Match	14.68;	Score 237.2;	DB 9;	Length 311;
Best Local Similarity	87.9%;	Pred. No. 4.2e-58;		
Matches 254; Conservative	0;	Mismatches 35;	Indels 0;	Gaps 0;

```

RESULT 9
US-09-938-842A-510
: Sequence 510, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Krebs, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: FILE REFERENCE: SAME, AND METHODS OF USE
: CURRENT APPLICATION NUMBER: US/09/938, 842A
: CURRENT FILING DATE: 2001-08-24

```

? PRIOR APPLICATION NUMBER: US 60/227,866
 ? PRIOR FILING DATE: 2000-08-24
 ? PRIOR APPLICATION NUMBER: US 60/264,647
 ? PRIOR FILING DATE: 2001-01-16
 ? PRIOR APPLICATION NUMBER: US 60/300,111
 ? PRIOR FILING DATE: 2001-06-22
 ? NUMBER OF SEQ ID NOS: 5379
 ? SEQ ID NO 510
 ? LENGTH: 1002
 ? TYPE: DNA
 ? ORGANISM: Arabidopsis thaliana
 ? US-09-938-842A-510

OY 933 GTCCTGAGATGGACATGATGCTGG 961
| | | | | | | | | | | | | | | | | |
DB 848 GCTCAGAGACGGAACGGTCACAAATATGG 876

RESULT 10

US-09-998-598-2482
; Sequence 2482, Application US/09998598
; Patent No. US20020150922A1

GENERAL INFORMATION:

; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2482
; LENGTH: 361
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-998-598-2482

Query Match 10.5%; Score 170.4; DB 10; Length 361;

Best Local Similarity 91.4%; Pred. No. 6.1e-39;

Matches 180; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 121 CCACAGCGCGCCGTGGTGGATGTCCTTCACAGCGCATCAGCCCTTACGCTACTTTC 180
| | | | | | | | | | | | | | | | | |
DB 1 CCACAGCGCACCCGTGGTGGATGTCCTTCACAGCGCATCAGCCCTTACGCTACTTTC 60
| | | | | | | | | | | | | | | | | |
OY 181 GATCAGCGCTTGAAGATGCGCAAGCCCATGCTCCGCCAGGAGATACAGAGACTGGAT 240
| | | | | | | | | | | | | | | | | |
DB 61 AATCAGCGCTTGAAGATGCGCAAGCCCATGCTCCGCCAGGAGATACAGAGACTGGAT 120
| | | | | | | | | | | | | | | | | |
OY 241 TGGAACTTTTGGTCAATTAAGTGTCTGTTGGGGTCAACATTTGAATTAAGATGGCAC 300
| | | | | | | | | | | | | | | | | |
DB 121 TGGAACTTTTGGTCAATTAAGTGTCTGTTGGGGTCAACATTTGAATTAAGATGGCAC 180
| | | | | | | | | | | | | | | | | |
OY 301 CAAAGCTGCGACAGCAG 317
| | | | | | | | | | | | | | | | | |
DB 181 CAAAGCAGCTACAGCAG 197
| | | | | | | | | | | | | | | | | |

RESULT 11

US-09-770-444-220
; Sequence 220, Application US/09770444
; Patent No. US20020023280A1

GENERAL INFORMATION:

; APPLICANT: Gorlach, Jörn
; APPLICANT: Ah, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maïa
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PAPA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 220
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana

US-09-770-444-220

Query Match 9.4%; Score 153.2; DB 10; Length 469;

Best Local Similarity 60.8%; Pred. No. 6.1e-34;

Matches 248; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

OY 120 GCCACAGCGCGCCGTGGTGGATGTCCTTCACAGCGCATCAGCCCTTACGCTACTTTC 179
| | | | | | | | | | | | | | | | | |
DB 5 GCGTCGAGACCTGTTGGATTTGTTTACAGTCCATCCTCCTGATGTTCTTCC 64
| | | | | | | | | | | | | | | | | |
OY 180 TGATCAGCGCTTGAAGATGCGCAAGCCCATGCTCCGCCAGGAGATACAGAGACTGGA 239
| | | | | | | | | | | | | | | | | |
DB 65 TCATCAGCGCAAGTAAAGATTCACACCAATGTTGAGAAATGGGAACTGGAGATTGGA 124
| | | | | | | | | | | | | | | | | |
OY 240 TTGAACATTTTGGTCAATTAAGTGTCTGTTGGGGTCAACATTTGAATTAAGATGCA 299
| | | | | | | | | | | | | | | | | |
DB 125 TTGGTCAATTTGAAGATGCGCAAGCCCATGCTCCGCCAGGAGATACAGAGACTGGA 184
| | | | | | | | | | | | | | | | | |
OY 300 CAAAGCTGCGACAGCAGCTGACACTTCACAGCCCAAGATTAAGATGCGCTCAGAG 359
| | | | | | | | | | | | | | | | | |
DB 185 TAGTGAGCTTGCATCTGCTGATTTTTCAGCAGAGCTTTGGAGTCTTGACGTGGG 244
| | | | | | | | | | | | | | | | | |
OY 360 ATGAATGATGACCTTGGCTCATTAAGACATGTCACAGATGAGATTTACACAGGATA 419
| | | | | | | | | | | | | | | | | |
DB 245 ATGCTTTGATTTTGGACACAGCATTAATGTTGAGCATGCGCTTCACAGGATA 304
| | | | | | | | | | | | | | | | | |
OY 420 GCAATTAAGTCAATGAGGAGACAGATTAAGTGTGCGCATATATGATGACCAAC 479
| | | | | | | | | | | | | | | | | |
DB 305 CGAATTAATCAATCAGAGGAGATTTGAGAAATTCGTCTGTTTGCATTTGATGCT 364
| | | | | | | | | | | | | | | | | |
OY 480 CTGAAGCAGACCTTAAGCAATCACTGCGCACACTTCTGTTATTAANA 527
| | | | | | | | | | | | | | | | | |
DB 365 TGGATGACCTCCTCCTACAGAAATGATTAATCTCTGTTCTACAGAA 412
| | | | | | | | | | | | | | | | | |

RESULT 12

US-09-969-708-323/c
; Sequence 323, Application US/09969708
; Patent No. US20020102532A1

GENERAL INFORMATION:

; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 323
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-969-708-323

Query Match 6.7%; Score 109.4; DB 10; Length 419;

Best Local Similarity 70.0%; Pred. No. 2e-21;
Matches 299; Conservative 0; Mismatches 96; Indels 32; Gaps 10;

QY 1226 TGAGGCAAGAGAGGCAAAATATTTGGGATATGATAGTCCAGCAGCAAGCT 1285
DB 418 TAAGGCAAGAGAGGCAAAATATTTGGGATATGATAGTCCAGCAGCAAGCT 359
QY 1286 -----ACTGAGTGTGGCCCGTGAAGAAATGGCTGAGTGTGAGTGCAGCAGCAGAG 1340
DB 358 AACTAAGTGTGGTACCTGTAAGTAAACCAAGTGC-----AGATGAAGAGGAGTGG 303
QY 1341 A-TTGGTGCACATAGTGGCAGTGGCTGTGG-----AATGAAGCAACT 1390
DB 302 ACTTATCCCTCTATATGATAGTGGCTGTGGCTTTTAAAGAAATATATACAGCAACA 243
QY 1391 TACAACTTCACAT-TTACACCTAAATTTCTTTAGCTTTTATGT---TATGAAGAA 1446
DB 242 TCCAAATTTCTATATTAACAATAGGCTCTGTAGCTTTATGTATATGAGAAAGAA 183
QY 1447 AAATATATGGCTATTTTCTGACTTCCCTTAAAGAAATGGCTTTTGTCCCTGCC 1506
DB 182 AACTATATTTGGCTGATTTT---TCTGATCTTAAAGCAAGATGCTTTCTTTTTCG 127
QY 1507 T-AGTATGAAGAGAGCAATATACATATTAAGTAAACGGTTTGAATCTTTGATGTA 1564
DB 126 TTAGCTTGAAGAGAGCAATATATATTAAGTAACTGTTGATTTTC--TGTTCAT 69
QY 1565 CAAGACTGCTTCAAGACAGTCA-TATTTTATTTATTAATTAATGCTTAAAT 1623
DB 68 TGTACACTGCTTCAAGACATTAATTTGTTTGTAGTGTCTAAATTAATGCTTAAAC 9
QY 1624 AAAAAA 1630
DB 8 AAAAAA 2

RESULT 13
US-09-924-035A-278/c
Sequence 278, Application US/09924035A
Patent No. US20020142319A1
GENERAL INFORMATION:
APPLICANT: Glaxo, Jn
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2011US
CURRENT APPLICATION NUMBER: US/09/924,035A
CURRENT FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/148,784
PRIOR FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 900
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 278
LENGTH: 463
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(463)
OTHER INFORMATION: n = A,T,C or G
US-09-924-035A-278

Query Match
Best Local Similarity 64.7%; Pred. No. 3.3e-11;
Matches 108; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 802 TGGGGTGAAGAGACTTAACTGTACAGATGATTTTAAACAGTGGAGAGACTTGA 861
DB 463 TGGTGTGAAGAGATATGTGGTTCGACTCTTTGATTTCCACCGGAGAGAGATGG 404
QY 862 ATCTCAAAAGCTGCTCCATTCATCTGTGATGATGATGATGATGATGATGATGAT 921
DB 403 ATGCAACAANNACATCATGATGATGATGATGATGATGATGATGATGATGATGATG 344

QY 922 CTATGCCAGCGTCTTGAAGATGGACATTTGATTTGGCAAACTG 968
DB 343 TTAGCTTCAAGGCTGTGAAGATGGTATCATCATGATTTGGCAAACTG 297

RESULT 14
US-09-923-876-3070
Sequence 3070, Application US/09923876
Patent No. US2002013958A1
GENERAL INFORMATION:
APPLICANT: Lalquid, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (Ito)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923,876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 3070
LENGTH: 262
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US2002013958A1 700161620H1
NAME/KEY: unsure
LOCATION: 45
OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-3070

Query Match
Best Local Similarity 58.6%; Pred. No. 6.6e-09;
Matches 112; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 798 TTGTGGGGTGAAGAGACTTAACTGTACAGATGATTTATTAACAGTGGAGAGAGT 857
DB 7 TTGTAGCTGGAGAGAGATTTGTGGTTCATGATTTCAATTTCTTCACTGTAAGAA 66
QY 858 TAGAATCTCAAAAGGCTCACTTTGTCCTCATTCAGTGTGAGTATGATGATGGG 917
DB 67 TACTGTGAACAAGGCAATATGACCGGTGACCTGCTGCTGCTGCTGCTGCTGCTG 126
QY 918 AACTATATCCAGCGGTTCTGAAGATGGAGATGAGATTTGGCAAACTGTGTAGGAA 977
DB 127 AATCATACCATCGGGTTCAGAAATGACCATTAAGATCTGGAGCTTACCCAGTTA 186
QY 978 AGACCTATGCG 988
DB 187 ATGCTGATGAC 197

RESULT 15
US-09-796-692-7155
Sequence 7155, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Galger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077,001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17

Accession	Gene	Protein	Species	Length (aa)	Weight (kDa)	PI	Ref
U00001	actA	Actin	Human	375	42	5.2	1
U00002	actB	Actin	Human	375	42	5.2	2
U00003	actC	Actin	Human	375	42	5.2	3
U00004	actD	Actin	Human	375	42	5.2	4
U00005	actE	Actin	Human	375	42	5.2	5
U00006	actF	Actin	Human	375	42	5.2	6
U00007	actG	Actin	Human	375	42	5.2	7
U00008	actH	Actin	Human	375	42	5.2	8
U00009	actI	Actin	Human	375	42	5.2	9
U00010	actJ	Actin	Human	375	42	5.2	10
U00011	actK	Actin	Human	375	42	5.2	11
U00012	actL	Actin	Human	375	42	5.2	12
U00013	actM	Actin	Human	375	42	5.2	13
U00014	actN	Actin	Human	375	42	5.2	14
U00015	actO	Actin	Human	375	42	5.2	15
U00016	actP	Actin	Human	375	42	5.2	16
U00017	actQ	Actin	Human	375	42	5.2	17
U00018	actR	Actin	Human	375	42	5.2	18
U00019	actS	Actin	Human	375	42	5.2	19
U00020	actT	Actin	Human	375	42	5.2	20
U00021	actU	Actin	Human	375	42	5.2	21
U00022	actV	Actin	Human	375	42	5.2	22
U00023	actW	Actin	Human	375	42	5.2	23
U00024	actX	Actin	Human	375	42	5.2	24
U00025	actY	Actin	Human	375	42	5.2	25
U00026	actZ	Actin	Human	375	42	5.2	26
U00027	actA	Actin	Human	375	42	5.2	27
U00028	actB	Actin	Human	375	42	5.2	28
U00029	actC	Actin	Human	375	42	5.2	29
U00030	actD	Actin	Human	375	42	5.2	30
U00031	actE	Actin	Human	375	42	5.2	31
U00032	actF	Actin	Human	375	42	5.2	32
U00033	actG	Actin	Human	375	42	5.2	33
U00034	actH	Actin	Human	375	42	5.2	34
U00035	actI	Actin	Human	375	42	5.2	35
U00036	actJ	Actin	Human	375	42	5.2	36
U00037	actK	Actin	Human	375	42	5.2	37
U00038	actL	Actin	Human	375	42	5.2	38
U00039	actM	Actin	Human	375	42	5.2	39
U00040	actN	Actin	Human	375	42	5.2	40
U00041	actO	Actin	Human	375	42	5.2	41
U00042	actP	Actin	Human	375	42	5.2	42
U00043	actQ	Actin	Human	375	42	5.2	43
U00044	actR	Actin	Human	375	42	5.2	44
U00045	actS	Actin	Human	375	42	5.2	45

XX 15-JUN-2000.
PF 10-DEC-1999; 99NM-U529267.
PR 10-DEC-1998; 98US-O111668.
PA (UYVA-) UNIV VANDERBILT.
XX
PI Datta PK, Moses HL;
DR WPI, 2000-442141/38.
XX P-PDB; AAB01250.

Pr Serine-threonine kinase receptors associated protein and the
Pt polynucleotide encoding the receptor, useful for treating disease
Cc associated with transforming growth factor beta activity such as cancer

Claim 18; Page 129-130; 134pp; English.

This sequence encodes the murine serine-threonine kinase receptor associated protein (STRAP). The STRAP polypeptide is capable of modulating transforming growth factor beta (TGF-beta) activity. TGF-beta belongs to a family of multifunctional cell regulatory factors, which bind to cell surface receptors. In certain cells TGF-beta promotes cell proliferation, and in others it suppresses proliferation. Some tumour cells use TGF-beta as an autocrine growth factor. STRAP enhances the anti-proliferative activity of TGF-beta. The invention includes isolated and purified anti-STRAP antibodies, and a hybridoma cell line producing the monoclonal antibody. STRAP exhibits cyostatic, antihypertensive, antidiabetic, and nephrotropic activity. The STRAP gene can be used to detect STRAP encoding RNA. Anti-STRAP antibodies are used to detect STRAP polypeptides in a sample. Compositions comprising STRAP polypeptides can be used to treat disorders associated with TGF-beta, such as cancer, atherosclerosis and diabetes.

Sequence 1630 BP; 455 A; 346 C; 406 G; 419 T; 4 other:

Query Match 99.8%; Score 1626; DB 21; Length 1630;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1627; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DY 1 TTATACCGCTACCAGGTGGGAGCCGTGAAGTGAGGCAGGCCGGCTTCGCGACTCCTCC 60
Db 1 TTTTACCGCTACCAGGTGGGAGCCGTGAAGTGAGGCAGGCCGGCTTCGCGACTCCTCC 60

OY 61 CCTCTCCCTCCCTCCCTCCCTCGCCCATGAGCACAGCGCCCTCACTTGTCTCGGG 120
Db 61 CCNCTCCCTCCCTCCCTCCCTCGCCCATGAGCACAGCGCCCTCACTTGTCTCGGG 120

OY 121 CCACAGCGGCCCCGTGGTAGATTGAGCCCTTACAGCGGCATCAGCCCTTAAGGCACTTTGT 180
Db 121 CCACAGCGGCCCCGTGGTAGATTGAGCCCTTACAGCGGCATCAGCCCTTAAGGCACTTTGT 180

OY 181 GATCAGCGCTTTGCAAAAGATGGACAAGCCATGCTCCGCCAAGAGATPACAGAGACTGGAT 240
Db 181 GATCAGCGCTTTGCAAAAGATGGACAAGCCATGCTCCGCCAAGAGATPACAGAGACTGGAT 240

OY 241 TGGAACATTTTTGGGCTCATTAAGAGTCTGCTTTGGGGTGCAAACATTGAATAAGATGCGAC 300
Db 241 TGGAACATTTTTGGGCTCATTAAGAGTCTGCTTTGGGGTGCAAACATTGAATAAGATGCGAC 300

OY 301 CAAGAAGCGGACACAGAGCTCAGACTTCACAGCCAAAAGTATGGAGTCCGGTCTCAGAGAGA 360
Db 301 CAAGAAGCGGACACAGAGCTCAGACTTCACAGCCAAAAGTATGGAGTCCGGTCTCAGAGAGA 360

OY 361 TGAATTGATGACCTGGCTCATTAAGACATTTGTCAAGCTGTGGATTTCACACAGATAG 420
Db 361 TGAATTGATGACCTGGCTCATTAAGACATTTGTCAAGCTGTGGATTTCACACAGATAG 420

OY 421 CAATTACCTGTTAACTGGGGACAGAGATAAACGTGCTCGGATATATAGACTTGAACAAC 480
Db 421 CAATTACCTGTTAACTGGGGACAGAGATAAACGTGCTCGGATATATAGACTTGAACAAC 480

Dy 421 CAATTACCTGTTAACTGGGGACAGAGATAAACGTGCTCGGATATATAGACTTGAACAAC 480

Qy	481	TGAGCACAACCTTAAGAAATACAGGGGCCACACTGCGATTTTAAAGGCTCTGGTG	540
Db	481	TGAACCAACCTTAAGAAATCAGGGCACCTCTGGATTTAAAGGCTCTGGTG	540
Qy	541	CAGTACGATTAACAGATCTCTTACGGCGATGATTAACCTGTCCTGGGATCATGC	600
Db	541	CAGTACGATTAACAGATCTCTTACGGCGATGATTAACCTGTCCTGGGATCATGC	600
Qy	601	CACAAATGACAGAGTGAANAATCTCTCAATTTATATGTCTGTATACACATCGATATAT	660
Db	601	CACAAATGACAGAGTGAANAATCTCTCAATTTATATGTCTGTATACACATCGATATAT	660
Qy	661	TCCCTAAGAGAGATTTTGTATTACTATATGAGACATCTATTTGCTTTCATATGTGAGT	720
Db	661	TCCCTAAGAGAGATTTTGTATTACTATATGAGACATCTATTTGCTTTCATATGTGAGT	720
Qy	721	AAGCTGAGCAATTAATATCCTTTGAGCTCTCTGCACATCATCTCGCTCTNTTCA	780
Db	721	AAGCTGAGCAATTAATATCCTTTGAGCTCTCTGCACATCATCTCGCTCTNTTCA	780
Qy	781	TCCAGAGAGAGGATTTCTGTGCGGGTGAGCAAGACTTTAACTGTACAACTATGATTA	840
Db	781	TCCAGAGAGAGGATTTCTGTGCGGGTGAGCAAGACTTTAACTGTACAACTATGATTA	840
Qy	841	TAACTATGGAGAAAGATTAGAAATCTTACAAAGTCACTTTGGTCCATTCACTGTGTAG	900
Db	841	TAACTATGGAGAAAGATTAGAAATCTTACAAAGTCACTTTGGTCCATTCACTGTGTAG	900
Qy	901	ATTCACTCCTGATGGGAACTCTATGCCAGCGGTTGTGAAGATGGACAATTGATGTAG	960
Db	901	ATTCACTCCTGATGGGAACTCTATGCCAGCGGTTGTGAAGATGGACAATTGATGTAG	960
Qy	961	GCAAACTGTGTGTAGGAAGACCTATGCGCTGTGGAATGCGTGNTTCTGAGGAGACAG	1020
Db	961	GCAAACTGTGTGTAGGAAGACCTATGCGCTGTGGAATGCGTGNTTCTGAGGAGACAG	1020
Qy	1021	CGGGAACTGTGCAAAAGCCAAAGATCGGATTTCCAGAAACGCAAGAGAAAGCTGGCAGA	1080
Db	1021	CGGGAACTGTGCAAAAGCCAAAGATCGGATTTCCAGAAACGCAAGAGAAAGCTGGCAGA	1080
Qy	1081	AGAAATGCTTTCAGAGAAATCAATTTCCATCTTATCATCAACTCTGAAGTTAAGCCTG	1140
Db	1081	AGAAATGCTTTCAGAGAAATCAATTTCCATCTTATCATCAACTCTGAAGTTAAGCCTG	1140
Qy	1141	AGCATCAGACGTGTGCTGCCGAAACCATATGTTCATGTAGGATTAACAACGACAGACAACA	1200
Db	1141	AGCATCAGACGTGTGCTGCCGAAACCATATGTTCATGTAGGATTAACAACGACAGACAACA	1200
Qy	1201	TCCGCTTTCAGAGTACTGTCTGCTCAGGCAAGAGAGGAGAAATTTGGGCGATATG	1260
Db	1201	TCCGCTTTCAGAGTACTGTCTGCTCAGGCAAGAGAGGAGAAATTTGGGCGATATG	1260
Qy	1261	AGTTAGCTCAGTGCACGAACAGCTACTCAGTGTGCGCGATGAAATATGGCTGAGTG	1320
Db	1261	AGTTAGCTCAGTGCACGAACAGCTACTCAGTGTGCGCGATGAAATATGGCTGAGTG	1320
Qy	1321	TCTATAGTGTGACGAGAGAGAGATTTGTGTCACATGTGCAATAGCCTGCTGTTTGAATGA	1380
Db	1321	TCTATAGTGTGACGAGAGAGAGATTTGTGTCACATGTGCAATAGCCTGCTGTTTGAATGA	1380
Qy	1381	AAAGCCAACTTACATCTCCATTTTACACCTAAATTTCTTTAGCTTTATGTATGAA	1440
Db	1381	AAAGCCAACTTACATCTCCATTTTACACCTAAATTTCTTTAGCTTTATGTATGAA	1440
Qy	1441	GAAAGAAAATATATGGCTATTTTCTGACTTTCCTTTAAAGAAATGCTTTTGTGCT	1500
Db	1441	GAAAGAAAATATATGGCTATTTTCTGACTTTCCTTTAAAGAAATGCTTTTGTGCT	1500
Qy	1501	CTTGCGCTGTATGAAGAGGAGAAATTCATGATTAAGTAACTACCGTTTGATCTCTTTCAT	1560
Db	1501	CTTGCGCTGTATGAAGAGGAGAAATTCATGATTAAGTAACTACCGTTTGATCTCTTTCAT	1560

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 33472; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 15320 BP; 4488 A; 2687 C; 3177 G; 4968 T; 0 other;
SQ
Query Match 49.9%; Score 812.8; DB 22; Length 15320;
Best Local Similarity 78.9%; Pred. No. 7.9e-224;
Matches 1179; Conservative 0; Mismatches 270; Indels 46; Gaps 16;
QY 163 GCCTACGGCTACTTCTGATCAGCGCTTGCAAGATGGCAAGCCATGCTCCGACAGG 222
DB 13046 GCTTATGATGATTTCTTAATCAGTGCCTGCAAAACGCAACCTATGCTATGCCAGG 12987
QY 223 ACATACAGAGACGATGATGGAACATTTGGGTCAATAAGGTGCTGGTGGGCGCAAC 282
DB 12986 AGATACAGAGACGATGATGGAACATTTGGGTCAATAAGGTGCTGGTGGGCGCAAC 12927
QY 283 ATTGAATAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 342
DB 12926 ACTAATAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12871
QY 343 GGATGCGGTCTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 402
DB 12870 GGATGCTATCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12811
QY 403 GGATGCTATCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 462
DB 12810 GGATGCTATCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12751
QY 463 ATATGACTTGAACCAACCTGAAAGCAGAACCTTAAGAAATGATGATGATGATGATGAT 522
DB 12750 ATATGACTTGAACCAACCTGAAAGCAGAACCTTAAGAAATGATGATGATGATGATGAT 12691
QY 523 TAAAAAGCCTCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
DB 12690 TAAAAAGCCTCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12631
QY 583 TCGG-CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 641
DB 12630 TAAGACTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12571
QY 642 TTAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
DB 12571 TTAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT

OY	642	TACACACATGGAGTATATTCGCGAAGGACAGATTTTGGCTATTAATCTATATGACACATCTA	701
Db	12581	CTAGTACTATGGAATATATTTCTGAGGAGAGATTTTGGTAATAACTTATGTGCTGATCCA	125222
OY	702	TTGCTTTTCATATGTGACATTAAGTCTGGAGCCAAATTAATCTTTGAAGTCTCCGACACA	761
Db	12521	GTCTTTTGGACATGACATGATTAAGTTGGAACCAATTAATCTTTCAAGCTCCTGCAACA	12462
OY	762	TCAATTCGTGGCTGCTTTCATCCAGAGAAAGAGAGTTCTTTGTGGGTGGAGAAAGCTTTA	821
Db	12461	TTTACCTGGAATCTCTTCACCTTGAGAAAGAAATTTCTTTGTCCAGGTGGTGAAGATGTTA	12402
OY	822	AACTTACACAGTATGATTTTAAACATGAGAGAAAGATTAAGATTCCTTCAAGATCTCTTG	881
Db	12401	AACCT-TATAAGTATGATGATTAATATGAGAAAGAAATTAATCTTCAGAGACACTCTG	12343
OY	882	GTCCCAATTCACCTGTGACATTCAGTCCCTGATGGGGAAGTCTATGCGACGGTCTGAA	941
Db	12342	GTCTTATTCAGTGTGGAGATTTGATCTCGATGAGAACATCTATCCAGTGGTCTGAA	12283
OY	942	ATGGGACATTTAGATATGTGGCAAACTGTGTGAGAAAGACATATGAGCTGTGGAATGCG	1001
Db	12282	ATGGACACTGTAGAC--TGGCAACGTGTGTAGAAACATATGCGCTTGTGGAATGTG	12225
OY	1002	TGTTTCTGTGAGAGACACGGGGAAACCTGGCAAAACCAAGATCGGATTTCCAGAAACG	1061
Db	12224	TGTTCTCTGAAGAAAGTACTGTTGAGCTGTGGCAACCCAAAGTTGTTTCCAAAGACAA	12165
OY	1062	CAGAGAGAGCTGGCAGAAAGAAATTTGCTTCAGAAATTCAGATTCATCTATTCATCAA	1121
Db	12164	GAGAAAC-----CGTAGAAAGAAATTTGGCTTCAGACATTCAGATTCATCTATTCCTCAA	12111
OY	1122	CTTCCTAAATTTAAGGCTTAGCATGACAGCTGTGGTGGCAACCAT--ATGTTATGG	1178
Db	12110	CTCTCAATTTAAGGCTGTACATATAGCATGTGTGCCAAGACAGATTAACATCGACTAG	12051
OY	1179	ACTAAACAAGCAGAGACAAACATCCGCCCTT-CAGATTTACTGTGTGCTCGAGCCAAAGAG	1237
Db	12050	AACGAGCAAGCAGAGAAAGAAAGCATTCAGCCCTTCAGAGTTACTCTGCTTAAGGCAAAAC	11991
OY	1238	GGCAGAAATTTGGGGCATATAGTATGCTTCACATGCTGACACACACT-ACTCAGTGTG	1296
Db	11990	ACCAATTAATATGACATATGAAATATGCTCCATGCTGGAACCAACTCAACATGCTGTA	11931
OY	1297	CCCTGAGTGAATAATGGCTGAGTGTCTGAGGTGAGCAGAGAGGATTTGCTGCATAGT	1356
Db	11930	CCCAATAGGAAATCTCAAGGTATCAGATG-AAGGGAGGTGAGTTATCTCTCTGTAGT	11872
OY	1357	GGCATAGGCTG--CTGTTGGATGAAAGAAAGCAACTTACAA-----TCTCCATTTT	1405
Db	11871	ACAAATGGCTGTCACTTTTAAATGAATATGTACAAAGTATCAACATCTCTCTTTATTA	11812
OY	1406	ACACCTAAATTTCTTTAAGCTGTTATG--TTATGAGAAAGAAATAATATGGCCTAT	1462
Db	11811	CAATTAAGATTTGTGCACTGTTTATATGCTATTAATGAGAAAGAAATAATATGGCCTAT	11752
OY	1463	TTTTCTGAC-TTTCCTTAAAGAAAGATGCTTTTGTCTCTGCTG-----TATGA	1515
Db	11751	TTTGTGACTTTTCCCTTAAAGCAAGATGTCTTTGTGCTATATGTTTGTGTTGGTTG	11692
OY	1516	AGAGAGAAATATCATGATTAAGTAAACCGTTTGATCTCTTTCAATTCACAGAGCTGCT	1575
Db	11691	TAAAGAGGGAATACATGATTAAGTAAAGTAAAGTAAATTTCTTCAATGTCAC-----ACTGG	11636
OY	1576	TCAGACAGCTCA-TATTTAGTATCTTAAATTAATATGGCTCTAAATAATAAAA	1629
Db	11635	TCCTGACATCTAATATTTTATTTAGTGTCTAATATTAATATGGCTCTGAAACAAAAA	11581

RESULT 8
AAC78126
ID AAC78126 standard; cDNA; 1430 BP

[illegible]

DB 362 CCGCAACCAATCAATTCGATCTTTCCTGAGAAAGATTTCTTGCGAGCGGNGAA 421
QY 812 GAAGACTTAAGTACAGTATGATTATA 842
DB 422 NA---TTTAACTTTTAAAGTATGATTATA 448

RESULT 11
AAC46123
ID AAC46123 standard; DNA; 1538 BP.

AAC46123:
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 48986.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126284.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134216.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0136229.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PN EPI033405-A2.
 XX 06-SEP-2000.
 PD
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128224.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 17-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140693.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 19-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 06-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148341.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.

20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158322.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0158294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 14.9%; Score 243.4; DB 21; Length 1540;
 Best Local Similarity 56.4%; Pred. No. 1.3e-59;
 Matches 495; Conservative 0; Mismatches 374; Indels 9; Gaps 2;

100 GAGCGCGCTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 159
 379 GACTCCGCTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 438
 160 CACGCGCTGCTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 219
 439 CACTCCGCTGCTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 498
 220 GCGAGATGCGAGAGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 279
 499 TGGGGAATGCGAGAGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 558
 280 AACATGATGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 339
 559 TTGCTGATGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 618
 340 ATGGGATGCGGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 399
 619 TTGGGATGCGGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 678
 400 TTGGGATGCGGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 459
 679 ATGGGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 738
 460 CATATATGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 519
 739 TGTTCGACTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 798

520 TATTAAGAGGCTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 573
 799 TATCAGAACACTGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 858
 574 TAAAGCTGCGGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 633
 859 TGGTGTGAGGCTGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 918
 634 TATGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 693
 919 GTCCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 978
 694 AGCATGCTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 753
 979 GTCAGCTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 1038
 754 TGGCAGCATGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 810
 1039 CTGCAATGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 1098
 811 AGAGACTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 870
 1099 AGAGACTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 1158
 871 AGCTGCTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 930
 1159 GGGACATGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 1218
 931 CGGCTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 968
 1219 AGGCTGCTGCTGCGAGAGCTGCGGCCACAGCGCGCGCTGCGATGCGCTTCAGCGGCAT 1256

RESULT 13

ID AAS58064 standard; cDNA; 313 BP.
 AAS58064
 AC AAS58064;
 DT 13-FEB-2002 (first entry)
 DE CDNA #740 encoding portion of a human colon tumour protein.
 KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
 OS Homo sapiens.
 PN WO200173027-A2.
 PD 04-OCT-2001.
 XX 22-MAR-2001; 2001WO-US09246.
 XX 24-MAR-2000; 2000US-191597P.
 PR 04-MAY-2000; 2000US-202024P.
 PR 05-MAY-2000; 2000US-202189P.
 XX (CORI-) CORIXA CORP.
 PA Meagher MJ, Xu J, King GE.
 PI WPL: 2001-611627/70.
 DR New colon tumour proteins and related nucleic acid, useful for
 PT treatment, prevention, diagnosis and monitoring of cancer -
 PT Claim 4; Page 174; 29pp; English.
 PS Th present invention relates to the isolation of novel cDNA sequences
 CC encoding for at least an immunogenic portion of human colon tumour
 CC proteins. The sequences of the invention are useful in pharmaceutical
 CC compositions and vaccines for the prevention and treatment of cancers
 CC such as colon cancer. They are also useful for the diagnosis and

1
2
3
4

100

100

100

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2003, 10:00:49 ; Search time 4360 Seconds

(without alignments)
10880.177 Million cell updates/sec

Title: US-09-856-836-1

Perfect score: 1630

Sequence: 1 ttaccgctaccgctaggga.....atgcctcttaataaaaaa 1630

Scoring table: IDENTITY_NTC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenBankl:*

1: gb_da:*

2: gb_hg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_ov:*

22: em_or:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hg_hum:*

31: em_hg_in:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pin:*

35: em_hg_rod:*

36: em_hg_mam:*

37: em_hg_vrt:*

38: em_sy:*

39: em_hggo_hum:*

40: em_hggo_mus:*

41: em_hggo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1626	99.8	1630	10 AF096285	AF096285 Mus muscu
2	1023.8	62.8	1867	9 HSM801659	AL136691 Homo sapi
3	1020	62.6	1712	9 BC000162	BC000162 Homo sapi
4	1020	62.6	1716	9 HA110025	AJ010025 Homo sapi
5	1018.4	62.5	1864	9 AB024327	AB024327 Homo sapi
6	1011.4	62.0	1820	6 AR053395	AR053395 Sequence
7	987.2	60.6	1563	9 AY049776	AY049776 Homo sapi
8	934.2	57.3	154115	9 AC010878	AC010878 Homo sapi
9	812.8	49.9	16680	9 AC010132	AC010132 Homo sapi
10	812.8	49.9	187017	2 AC010101	AC010101 Homo sapi
11	674.2	41.4	226520	10 AL645802	AL645802 Mouse DNA
12	665	40.8	73038	2 AC020884	Continuation (4 of
13	631.8	38.8	170393	10 AC109605	AC109605 Mus muscu
14	538	33.0	1829	5 AF315726	AF315726 Carassius
15	338	20.7	1591	3 AY061000	AY061000 Drosophill
16	285.6	17.5	413	9 HSU56430	U56430 Human Hela
17	243	14.9	313	6 AX261089	AX261089 Sequence
18	240.8	14.8	686	6 AR144171	AR144171 Sequence
19	240.8	14.8	686	6 AR176358	AR176358 Sequence
20	240.8	14.8	686	6 AX365643	AX365643 Sequence
21	237.2	14.6	311	6 AX393751	AX393751 Sequence
22	213.6	13.1	1411	8 AY084604	AY084604 Arabidops
23	212	13.0	1399	2 AY074392	AY074392 Arabidops
24	204.6	12.6	31772	8 AC020469	AC020469 Drosophill
25	204.6	12.6	188272	3 AC005639	AC005639 Drosophill
26	204.6	12.6	218565	3 AC009257	AC009257 Drosophill
27	178	10.9	182247	9 AC022073	AC022073 Homo sapi
28	170.8	10.5	223	6 A74365	A74365 Sequence 51
29	170.8	10.5	223	6 A77344	A77344 Sequence 51
30	134.2	8.2	60376	2 AC097388	AC097388 Rattus no
31	134.2	8.2	169695	2 AC106673	AC106673 Rattus no
32	125.8	7.7	169695	2 AC106673	AC106673 Rattus no
33	125.2	7.7	51381	2 AF322456	AF322456 Homo sapi
34	109.4	6.7	419	6 AX337285	AX337285 Sequence
35	89.6	5.5	89765	3 AC017970	AC017970 Drosophill
36	89.6	5.5	188873	3 AC099009	AC099009 Drosophill
37	89.6	5.5	196337	3 AC005894	AC005894 Drosophill
38	89.6	5.5	242172	3 AE003832	AE003832 Drosophill
39	83.6	5.1	2781	10 AF020683	AF020683 Rattus no
40	82.2	5.0	34410	3 AC008242	AC008242 Leishmani
41	82.2	5.0	116000	2 AC129714	AC129714 Leishmani
42	75.2	4.6	98151	2 AC119404	AC119404 Magnapor
43	66	4.0	10048	3 AY061931	AY061931 Drosophill
44	64.6	4.0	81370	8 AB017071	AB017071 Arabidops
45	64.6	4.0	117737	8 F6D8	AC008016 Arabidops

ALIGNMENTS

RESULT 1
AF096285
LOCUS
DEFINITION Mus musculus serine-threonine kinase receptor-associated protein
ACCESSION AF096285
VERSION AF096285.1 GI:4063382
KEYWORDS
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Datta,P.K., Chytil,A., Gorska,A.E. and Moses,H.L.
REFERENCE 1 (bases 1 to 1630)
AUTHORS
TITLE Identification of STRAP, a novel WD domain protein in transforming

Db 1621 AATTAATAAAA 1630

|||||

RESULT 2
HSM801659
LOCUS
DEFINITION
HSM801659 1867 bp mRNA linear PRI 10-MAR-2001
Homo sapiens mRNA; cDNA DKFZp564N1778 (from clone DKFZp564N1778);
complete cds.
AL136691
AL136691.1 GI:6807652

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1867)
Wieman, S., Well, B., Wellenreuther, R., Gassenhuber, J., Glasl, S.,
Ansoerg, M., Boecher, M., Blocher, H., Bauersachs, S., Blum, H.,
Laufer, J., Duesterhoeft, A., Beyer, A., Koehler, K., Strack, N.,
Mewes, H.W., Ottenwaelder, B., Obermaier, B., Tampe, J., Heubner, D.,
Wambolt, R., Korn, B., Klein, M., and Poustka, A.
Toward a Catalog of Human Genes and Proteins: Sequencing and
Analysis of 500 Novel Complete Protein Coding Human cDNAs
Genome Res. 11 (3), 422-435 (2001)
11230166
2 (bases 1 to 1867)
Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J., and
Wieman, S.
Direct Submission
Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wieman, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wieman@dkfz-heidelberg.de;
Sequencing consortium of the German Genome Project. This clone
(DKFZp564N1778) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
Location/Qualifiers
1. 1867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp564N1778"
/tissue_type="brain"
/clone_id="564 (synonym: hfbz2). Vector pAMP1; host
XL-2blue; sites NotI + SalI"
/dev_stage="fetal"
297..1349
/gene="DKFZp564N1778"
297..1349
/gene="DKFZp564N1778"
/note="un-Interacting protein"
/product="hypothetical protein"
/protein_id="CA66626.1"
/db_xref="GI:6807653"
/db_xref="sptrembl:O9NTR0"
/translation="MAMROTPLTSGHTRPVVDLAFSGITPYGYLISACKDKPKMLR
QGDPTGMITGFLGHKGAIVGATLNDKATAAATAADPTKAVDVAAGDELMLTANHI
VTVDTQDSNLTGCGDKLRIYDLNPEAPKPKISGHTSGIKALKASDKOLIS
ADKTVRLMDHATMTVEKSLNFMVSWEYIPEGHLLVITVGRSLAPFASVLDPIK
SEAPATINSASLHPEKELVAGGEDFKYKYVDYNGSELSEYKGFPGHIVRVSPP
GELIYSGSEDTLRIMQTVVKTGYIMKCVLPEBDSGELAKRIGPPTTELELEIA
SENSEIPEPSADPKA"
1829..1834
polya_signal
polya_site 1842..1834
BASE COUNT 532 a 388 c 422 g 525 t
ORIGIN
Query Match 62.8%; Score 1023.8; DB 9; Length 1867;

Best Local Similarity 82.5%; Pred. No. 1.8e-263;
Matches 1310; Conservative 0; Mismatches 246; Indels 31; Gaps 11;

QY 66 TCCCTCCCTCCCTGTCGCCATGATGAGCAGACGCGGCTCACTTCGCGGCCACA 125
|||
Db 277 TCGCTGCTGCTCGCGCGCGCATGGCATAGACACACGCGCTCACTGCTGCGCCA 336
QY 126 CGCGGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 185
|||
Db 337 CGCGGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 396
QY 186 GCGCTTGCAGAAAGCCATGAGCAGACGCGGCTCACTTCGCGGCCACA 245
|||
Db 397 GCGCTTGCAGAAAGCCATGAGCAGACGCGGCTCACTTCGCGGCCACA 456
QY 246 CATTTTGGGTCATTAAGAGTGTGTTGGGTCGCAACATTAAGAGTGTGTTGG 305
|||
Db 457 CATTTTGGGTCATTAAGAGTGTGTTGGGTCGCAACATTAAGAGTGTGTTGG 516
QY 306 CTGGCAGCAGCAGCTGACACTTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 365
|||
Db 517 CAGCTACAGCAGCTGACACTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 576
QY 366 TGATGACCTGCTCATTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 425
|||
Db 577 TGATGACCTGCTCATTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 636
QY 426 ACCTGNTAACCTGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 485
|||
Db 637 ATTTGTTAACCGGGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 696
QY 486 CAGAACCTTAAGCAATAGTGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGC 545
|||
Db 697 CAGAACCTTAAGCAATAGTGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGC 756
QY 546 ACGATTAACAGATCTTTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 605
|||
Db 757 AGGATTAACAGATCTTTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 816
QY 606 TGAAGAGGAAATCTCTGAATTTTATATGCTGTGATGACAGCAGCAGCAG 665
|||
Db 817 TGAAGAGGAAATCTCTGAATTTTATATGCTGTGATGACAGCAGCAGCAG 876
QY 666 AAGGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 725
|||
Db 877 AAGGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 936
QY 726 TGAAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 785
|||
Db 937 TGAAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 996
QY 786 AGAAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 845
|||
Db 997 AGAAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 1056
QY 846 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 905
|||
Db 1057 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 1116
QY 906 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 965
|||
Db 1117 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 1176
QY 966 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 1025
|||
Db 1177 GTGAGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTA 1236
QY 1026 AACTGGCAAGCAAGATGATGATTTCCAGAAACAGCAGAGAGAGAGAGAG 1085
|||
Db 1237 AACTGGCAAGCAAGATGATGATTTCCAGAAACAGCAGAGAGAGAGAGAG 1293
QY 1086 TTGCTTCAAGAGATTCATTCATTCATTCATTCATTCATTCATTCATTC 1145
|||

Db 666 TGACAGAGAGTAATCTCTAAATTTATATCTGTTACTAGTATGGAATATATTCCTG 725
 QY 666 AAGAGAGATTTGGTTATATCTATGAGAGATCTATGCTTTTCAATAGTCAAGTATG 725
 Db 726 AAGAGAGATTTGGTTATATCTATGAGAGATCTATGCTTTTCAATAGTCAAGTATG 785
 QY 726 TGAGAGCAATTAATCTCTGTTGAGAGATCTATGCTTTTCAATAGTCAAGTATG 785
 Db 786 TGAGAGCAATTAATCTCTGTTGAGAGATCTATGCTTTTCAATAGTCAAGTATG 845
 QY 786 AGAAGAGATTTCTGTTGAGAGATCTATGCTTTTCAATAGTCAAGTATG 845
 Db 846 AGAAGAGATTTCTGTTGAGAGATCTATGCTTTTCAATAGTCAAGTATG 905
 QY 846 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 905
 Db 906 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 965
 QY 906 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 965
 Db 966 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1025
 QY 966 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1025
 Db 1026 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1085
 QY 1026 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1085
 Db 1086 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1142
 QY 1086 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1142
 Db 1143 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1202
 QY 1143 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1202
 Db 1146 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1262
 QY 1146 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1262
 Db 1203 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1262
 QY 1203 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1262
 Db 1204 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1322
 QY 1204 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1322
 Db 1263 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1378
 QY 1263 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1378
 Db 1378 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1438
 QY 1378 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1438
 Db 1379 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1498
 QY 1379 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1498
 Db 1439 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1554
 QY 1439 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1554
 Db 1498 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1614
 QY 1498 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1614
 Db 1614 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1672
 QY 1614 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1672
 Db 1672 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1702
 QY 1672 GTGAGAGATTTGAGATCTATGCTTTTCAATAGTCAAGTATG 1702

LOCUS HAU10025 1716 bp mRNA linear PRI 12-MAR-1999
 DEFINITION Homo sapiens mRNA for unr-interacting protein.
 ACCESSION AJ010025.1 GI:4468217
 VERSION 1
 KEYWORDS RNA-binding protein; unr-interacting protein; unr1p gene.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1716)
 AUTHORS Hunt,S.L.
 TITLE Direct Submission
 JOURNAL Submitted (06-AUG-1998) Hunt S.L., University of Cambridge, Department of Biochemistry, 80 Tennis Court Road, Old Addenbrooke's Site, Cambridge, CB2 1QA, U.K.
 REFERENCE 2 (bases 1 to 1716)
 AUTHORS Hunt,S.L., Hsuan,J.J., Totty,N. and Jackson,R.J.
 TITLE unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA
 JOURNAL Genes Dev. 13 (4), 437-448 (1999)
 MEDLINE 99158740
 PUBMED 10049359
 FEATURES
 source location/Qualifiers
 1..1716
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="Hela cell"
 148..1200
 /gene="unr1p"
 148..1200
 /gene="unr1p"
 /function="Interacts with human cold shock domain protein, unr"
 /codon_start=1
 /evidence=experimental
 /product="unr-interacting protein"
 /protein_id="CAB38041.1"
 /db_xref="GI:4468218"
 /translation="MAMROTPLTSGHTRPVVDIAFSGITPYGYLISACKDKRMLR
 OGDPMGICFPLGKAVVATLINDKAAADFAKVDVAVSGELMTLARKHI
 VKYVDTOSSNLTGQDKLRIYDNLPEAPEKELISGTEGIRKALWCSDKQILS
 ADKTVRLMDHATMEVKSINFSVSMSEYIPDEGILVYTKRSIAFSAVADPIK
 SEPAPTNISASLHPEKELVAGGDFKLYKDYVSGEELSEYKHPGPIHCVRESPI
 GELYSAGSEDTLRIMQIVYVGYGMKVLPEEDSGELAKRKIFPETTELEETIA
 SENSDFIPSPADVKA"
 BASE COUNT 519 a 320 c 395 g 482 t
 ORIGIN
 Query Match 62.6%; Score 1020; DB 9; Length 1716;
 Best Local Similarity 83.2%; Pred. No. 1.9e-262;
 Matches 1323; Conservative 0; Mismatches 229; Indels 38; Gaps 13;
 QY 66 TCCCTCCCTCCCTGTCCTCCCATGATGAGCAGAGCCGCTCACTCTCGGGCACA 125
 Db 128 TCCCTGCTTGGCGCGCCATGAGCAGAGCCGCTCACTCTCGGGCACA 187
 QY 126 CGCGGCCGCTGTGATNTGGCTTCAGCGGATCAGCGCTTACGGCTCTTGATGA 185
 Db 188 CGCGGCCGCTGTGATNTGGCTTCAGCGGATCAGCGCTTACGGCTCTTGATGA 247
 QY 186 CGCGCTTGCAAAAGTGGCAAGCCCATGCTCCGCGAGGAGATACAGAGACTGATTGGA 245
 Db 248 CGCGCTTGCAAAAGTGGTAACCTATGCTACGCGAGGAGATACAGAGACTGATTGGA 307
 QY 246 CATTTTGGGTATAAAGTGTGTTGGGTGCAAACTGAATAAGATGCCACCAAG 305
 Db 308 CATTTTGGGTATAAAGTGTGTTGGGTGCAAACTGAATAAGATGCCACCAAG 367
 QY 306 CTGCGACAGCAGCTCAGACTTACAGCCAAAGTATGGATCGGCTCAGAGATGAAT 365
 Db 368 CAGCTACACAGCTCAGACTTACAGCCAAAGTATGGATCGGCTCAGAGATGAAT 427

Oy	366	TGATGACCTGGCTCATTAAGCAACATTTGCAAGAGCTGGATTTTCAACAGAGATTACAA	425
Oy	367	TT	426
Db	428	TGATGACCTGGCTCATTAAGCAACATTTGCAAGAGCTGGATTTTCAACAGAGATTACAA	487
Oy	426	ACCTGTAATCTGGGGGACAGGATAAAGCTGCGGCATATATGACTTTGAACAAACCTGAAG	485
Db	488	ATTTTGTAAACGGGGGACAGGATAAAGCTGTACCATATATGACTTTGAACAAACCTGAAG	547
Oy	486	CAGAACCTTAAGGAATTCAGTGGCCACACTTCCTGATATAAAAGGCTGGTGGAGAG	545
Db	548	CAGAACCTTAAGGAATATGATGGTCACTACTTGGTATAAAAAAAAGCTCTGGTGGAGAG	607
Oy	546	ACGATTAACAGATCCCTTTCAGGGATGATAAAGCTGTGGCTCTGGAGCATGCCAA	605
Db	608	AGGATTAACAGATCTTTCTTCTGCTGATGACAAAGCTGTGCACTTGGAGCATGCTACTA	667
Oy	606	TGACAGAGTGAATTCCTGAAATTTAATATGCTCTTTAGACAGATGAGGATATATCCG	665
Db	668	TGACAGAGTGAATTCCTGAAATTTAATATGCTCTTTAGATAGATGAAATATATTCG	727
Oy	666	AAGGAGAGATTTTGGTTATCTTATGAGCATCATATGCTTTTCAATGAGCAGTAAGTC	725
Db	728	AGGAGAGATTTTGGTTATCTTATGAGCATCATATGCTTTTCAATGAGCAGTAAGTT	787
Oy	726	TGAGCCAAATTAATCCCTTTGAAGCTCCGAGACCATCAATTCGCGTCTNTTCAATCCAG	785
Db	788	TGAGCCAAATTAATCCCTTTGAAGCTCCGAGACCATCAATTCGCAATCCCTTCAATCCG	847
Oy	786	AGAGAGAGTTTCTGTTTGGGGGAGAGAAAGCTTTAAAGCTGTACAAATATATATTAACA	845
Db	848	AGAGAGAGTTTCTGTTTGGGGGAGAGAAAGCTTTAAAGCTTTAATAGATATATATATA	907
Oy	846	GTGAGAGAAGATTAGAAATCCATCAAAAGTCACTTTGGTCCATTCAGTGTGTAGATATCA	905
Db	908	GTGAGAGAAGATTAGAAATCCATCAAAAGGCACTTTGGTCCATTCAGTGTGTAGATTTA	967
Oy	906	GTCCTGATGGGGAACCTATGACGAGGATTCGAGATGGAGACATATGAGATTGGGCAAA	965
Db	968	GTCCTGATGGGGAACCTATGACGAGGATTCGAGAGATGGAGACATATGAGATATGGCAA	1027
Oy	966	CTGTGTAGAGAAAGACCTATGCGCTGTGGAATTCGCTGTTCTGTAGAGAGACAGCGGG	1025
Db	1028	CTGTGTAGAGAAAGACCTATGCGCTTTGGAAATTCGTCTTCGTGAAAGATATGTTG	1087
Oy	1026	AACGCGAAAGCCAAAGATGCGATTTCCAGAAACAGCAGAGAAAGAGTGTGCACAAAGAA	1085
Db	1088	AGCTGCGAAAGCCAAAGATGTTGTTTCCAGAGACAGCAGAGAGAGCT---AABAAGAA	1144
Oy	1086	TTGCTTCAGAGAAATTCAGATTCATCTATTCATCACTCTGAAGTTAAGGCTGAGCAT	1145
Db	1145	TTGCTTCAGAGAAATTCAGATTTGCAATTCCTTCCTGATGTTAAGGCTCGAGAGCT	1204
Oy	1146	CAGACGTGTGCTGCCGAACCAT--ATGTTCAATGAGCTAATAACAGACAGCAAGCATCC	1203
Db	1205	CAGATCATTTGTTGCAAGTTAGTATTAACAATGACTATAAAACAGCAACAGAGAAAGCATCA	1264
Oy	1204	GCGCTT--CAGAGTTTACTGTCTGCGCCGAGAGCAAGAGGGCAGAAATATTTGGGCATATAG	1262
Db	1265	GCGCTTCCAGATTTACTGTCTGCTTTAAGGCGAAGAACAGCATTAATATGAGGAAAAATGAA	1324
Oy	1263	TTAGCTCCAGTGCAGCAAGCAAGCT---ACTCAGTGTCCCGGTGAGTGAATATGGCTGGA	1317
Db	1325	TTAGCTCCAGTGCAGTGAACAATCACTCACTCACTTGTGCTGTAAGGTAAAA---CTC	1380
Oy	1318	GTGTCTGAGAGTGCAGGAGAGGGA--TTGTGCTCACTAGAGCAATAGCCTGCTGTTT---	1373
Db	1381	AGGTCTCAGATGAGGGGGGGGGAGTTATCTCTTTATATAGCAGTGGCCTGTATCTTTT	1440
Oy	1374	-----GGAATGAAAGCAACTTACAATCTCCATT--TTACACTTAATTTCTTTTACT	1426
Db	1441	TATGATAATATATACAGGCAACATCCAAATTTCTATTATTAACAATTTAGGCTGTCTGTACT	1500
Oy	1427	GTTTATGT---TATGAGAGAGAAAAATATATTTGGCTATTTTTCTGACTTCCCTTAAG	1483

Db	Accession	Length	Score	DB	Gap	Indel	Mismatch	Conservative	Similarity	Match
Db	1501	1556	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1484	1541	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1557	1616	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1542	1600	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1617	1674	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675	1704	1018.4	9	262	38	230	0	83.1%	Best Match
QY	1601	1650	1018.4	9	262	38	230	0	83.1%	Best Match
Db	1675									

QY	261	AAGTGTGCTTTGGGGGTGCACAACTTGGTAAAGGATGGCCACCAAAAGCTGGCAGCAGCTG	320
Db	434	AAGGTGCTGTTGGGGGTGCACAACTGAATTAAGGATGCCACCAAAAGCAGCTACAGCAGCTG	493
QY	321	CAGATTTCACAGCCCAAAATATATGGGATGGGGCTCAGAGATGAATATGTAACCTGGCTC	380
Db	494	CAGATTTCACAGCCCAAAAGTGGGATGGCTGCTCAGAGATGAATATGTAACCTGGCTC	553
QY	381	ATAGCACATTGTCAAGACTGTGGATTTCCACACAGGATATAGCAATTAAGCTGTAATCTGGGG	440
Db	554	ATMAACACTTGTCAAGACTGTGGATTTCCACAGAGATATGAATATTTGTTAACCGGGG	613
QY	441	GACAGATTAACCTGCTGGCAGCTATATGACTTTGAACAAACCTGGAAACAGAACCTTAAGAAA	500
Db	614	GACAGATTAACCTGTTAGCCATATATGACTTTGAACAAACCTGGAAACAGAACCTTAAGAAA	673
QY	501	TCAGTGGCCACACTCTGTGATTTAAAAAGCTCTGTGGCAGTACGATTAACGATATCC	560
Db	674	TTAGTGGCTACTCTGTGATTTAAAAAGCTCTGTGGCAGTACGATTAACGATATCC	733
QY	561	TTTCAGCGGATATATAAAGTGTGGGCTGGGATCATCCCAATGAGCAGAAAGAAAT	620
Db	734	TTTCGCTATGCAAAAGCTGTTCGACTTTGGGATCATCTCTATGACAGAAAGAAAT	793
QY	621	CTCTGAATTTTAAATATGCTGTGTTAGCAGCATGGAGTATATTCCTGAAGAGAGATTTGG	680
Db	794	CTCTAAATTTTAAATATGCTGTGTTAGTATGGAATATATTCCTGGAGGAGAGATTTGG	853
QY	681	TTATTTACTATGAGCATCTATGTGCTTTTCATATGCAAGCAAGTGAAGTGGCCCAATTAAT	740
Db	854	TTTATTAACCTATGAGCATCTATGTGCTTTTCATATGTCAGATGAAGTTGGACCCCAATTAAT	913
QY	741	CCTTGAAGCTCCTGGACCATCAATTCGCGCTTTTCATCCAGAAAGAGATTTCTTG	800
Db	914	CCTTGAAGTCTCTGCACCATCAATTTGACATTCCTGATCCTGAGAAAGATTTCTTG	973
QY	801	TTGCGGGTGCAGAAAGCTTTAACTGTACAACTATATATATAAGTGGAGAAAGATTTG	860
Db	974	TTGCGCGGTTGAAGATTTTAACTTTATATAGTATATTTATATGTGGAGAAATTTG	1033
QY	861	AATCCTCAAAAGTTCACCTTGTGCCATTCACACTGTGTGAGATTCAGTCTGATGGGAAAC	920
Db	1034	AATCCTCAAAAGGACACTTTGGTCCATTCACAGTGTGAGATTTAGTCCGATGGGAAAC	1093
QY	921	TCATGCCACGCTTCTGAAAGTGGGACATTGAGATTTGGGCAACTGTGTGGGAAAG	980
Db	1094	TCATGCCACGTGTCTGAAAGATGGAACTATGAGACTATGGCAAACTGTGTGGGAAAA	1153
QY	981	CCATGAGCCTGTGGAAATGCGTGTTCCTCAGAGAACACACGGGGAACTGGCAAGCCAA	1040
Db	1154	CGTATGCGCTTTGAAATGTGTCTTCTTAAGAAATATGTGTGAGCTGGCAAGCCAA	1213
QY	1041	AGATCGGATTTCCAGAAACAGCAGAGAGAGAGCTGGCAGAAAGAAATTTGCTTCAGAGAA	1100
Db	1214	AGATTTGCTTTCCAGAGAACACAGAAAGAGAGAGCT--AGAAAGAAATTTGCTTCAGAGAA	1270
QY	1101	CAGATTCCATCTATTCATCAACCTCTGAAGTTTAAGGCTCAGAGCTGAGTGTGCTGC	1160
Db	1271	CAGATTCCATTTTCCCTTCAAGCTCTCTATATGTTAAAGGCTCAGAGCTCAATATGTTGCA	1330
QY	1161	GAACCAAT--ATGTTCAATGCACTTAACAGAGAGAGACAGACATCCGCTT--CAGAGTTAC	1217
Db	1331	GTTAGTATACAACTGACATTAACAAAGCAGAGAGAAAGATCAGGCTTCCAGAGTTAC	1390
QY	1218	TGTTGCTGCTGAGGCAAAAGAGGGCAGAAATTTGGGGCATATATGAGTTAGCTCCAGTGCAC	1277
Db	1391	TGTTGCTGCTTAAGGCAAAAGCAGATTAATATATGAGGAAATGAAATTAAGCTCCAGTCTG	1450
QY	1278	GAAACAGCT----ACTCAGTGTTCACCGTGAATGAATAATGGCTCAGTCAAGTGTCCAG	1332
Db	1451	GAAACACTTAATTAAGTGTGTTCAGCTGTGAAGTGAATA--CTCAGTGTCAAGATGAAG	1506
QY	1333	GCAGAGACA--TTGTGCTCACAATAGTGCATATAGCCTGCTGTTT-----GGAAATGAAA	1382

[illegible]

	LOCUS	AY049776	RESULT 7	
	DEFINITION	Homo sapiens unr-interacting protein (UNRIP) mRNA,	complete cds.	PRI 23-OCT-2001
	ACCESSION	AY049776		
	VERSION	AY049776.1	GI:16356636	
	KEYWORDS	.		
	SOURCE	Homo sapiens.		
	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	REFERENCE	1 (bases 1 to 1563)		
	AUTHORS	Liu,J., Zhou,Y., Peng,X., Yuan,J. and Qiang,B.		
	TITLE	Direct Submission		
	JOURNAL	Submitted (31-JUL-2001) Department of Biochemistry, Institute of Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong Dan San Tiao, Beijing 100005, P.R. China		
	FEATURES	Location/Qualifiers		
	source	1..1563		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
	gene	1..1563		
		/gene="UNRIP"		
	CDS	48..1100		
		/gene="UNRIP"		
		/codon_start=1		
		/product="unr-interacting protein"		
		/protein_id="AA15433.1"		
		/db_xref="GI:16356637"		
		/translation="MAMROTPLTCSGHTRPVVDIAFSGITPYGFILISACKDKPMLR OQDIDWIGTFLGHKGAWGATLNNDATKAATAADFTAVKWDVSDDELDTLNAKHIT VTVPFDOSNYNLITGGODKLRLTYDLNKPPEAPRELSGHTSGIKALWCSEQIJS ADDKVRLMDHATMETEVKSFLANMNSMSNEYIEGEELIVITYGSIFHSVASDPPIRK SEPAPRTINSASLTHPKEFLVAGGEDFLTYDVNSGBELSYSGHFPGPIHCVFSPSD GELYSGSEDGTLRLLMQTVYVKTYGLMKVCVLPEDDSGLAKPKIGPETTEELEELINA SENSDICEPSAPDVKA"		
	BASE COUNT	465 a 284 c 362 g 452 t		
	ORIGIN			
	Query Match	60.6%; Score 987.2; DB 9;	Length 1563;	
	Best Local Similarity	83.2%; Pred. No. 1.2e-253;		
	Matches 1274; Conservative	0; Mismatches 222;	Indels 36; Gaps 12;	
Dy	66	TCCTCCTCCCTCGTCTCGCGCATGAGGCATGAGCGCGCTCACTTGCTGGGCCACA	125	
Dd	28	TCGCTGCGCTTCGCGCGCATGAGCATGAGACGCGCTCACACTGCTGGGCCACA	87	
Oy	126	CGCGCGCCCTTGTTGAANTGGCTTAGCGGCATCAAGCCTTAGCGCTACTTTGATCA	185	

Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@atson.wustl.edu
 ----- Summary Statistics -----
 Center project name: H_NH0230E20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dpcpac.med.buffalo.edu>)

VECTOR: pBACe3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-278G12, 200 bp overlap;
 the clone sequenced to the left is RP11-92F3, 200 bp overlap.
 Actual start of this clone is at base position 1 of RP11-230E20;
 actual end is at base position 23434 of RP11-278G12.

RP11-230E20 contains polymorphisms with RP11-278G12.

Location/Qualifiers

```

FEATURES
source
    1..154115
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="2"
        /map="2"
        /clone="RP11-230E20"
        /clone_lib="RPc1-11"
        1..97
            /rpt_family="L1"
            158..389
                /rpt_family="MIR"
                450..788
                    /rpt_family="L1"
                    1032..1326
                        /rpt_family="Alu"
                        1744..1781
                            /rpt_family="MER1_type"
                            1793..2026
                                /rpt_family="Alu"
                                2032..2132
                                    /rpt_family="MER1_type"
                                    2164..2252
                                        /rpt_family="L1"
                                        2258..2613
                                            /rpt_family="L1"
                                            2633..2895
                                                /rpt_family="Alu"
                                                2899..2919
                                                    /rpt_family="TAA)n"
                                                    2923..3480

```

```

repeat_region
    3552..3659
        /rpt_family="L1"
repeat_region
    3670..3777
        /rpt_family="L1"
repeat_region
    4243..4534
        /rpt_family="Alu"
repeat_region
    4827..5132
        /rpt_family="Alu"
repeat_region
    5718..5925
        /rpt_family="MER1_type"
repeat_region
    6210..6792
        /rpt_family="L2"
repeat_region
    6933..6994
        /rpt_family="ERV1"
repeat_region
    6995..7535
        /rpt_family="ERV1"
repeat_region
    7516..7619
        /rpt_family="ERV1"
repeat_region
    7620..7921
        /rpt_family="Alu"
repeat_region
    7922..8073
        /rpt_family="ERV1"
repeat_region
    8078..8386
        /rpt_family="Alu"
repeat_region
    8436..8893
        /rpt_family="ERV1"
repeat_region
    9675..9781
        /rpt_family="MALR"
repeat_region
    10536..10774
        /rpt_family="MALR"
repeat_region
    10784..11233
        /rpt_family="MALR"
repeat_region
    11428..11544
        /rpt_family="MALR"
repeat_region
    11545..11899
        /rpt_family="MALR"
repeat_region
    11900..11920
        /rpt_family="MALR"
repeat_region
    12078..12359
        /rpt_family="Alu"
repeat_region
    12420..12706
        /rpt_family="MALR"
repeat_region
    14365..14403
        /rpt_family="AT-rich"
repeat_region
    14399..14928
        /rpt_family="MER1_type"
repeat_region
    14930..15505
        /rpt_family="ERV1"
repeat_region
    15849..16157
        /rpt_family="Alu"
repeat_region
    17559..17692
        /rpt_family="MIR"
repeat_region
    18143..18624
        /rpt_family="ERV1"
repeat_region
    18625..19034
        /rpt_family="ERV1"
repeat_region
    19035..19080
        /rpt_family="ERV1"
repeat_region
    19614..19860
        /rpt_family="MIR"
repeat_region
    20803..20906
        /rpt_family="MER1_type"
repeat_region
    20994..21169
        /rpt_family="L1"
repeat_region
    21214..21402
        /rpt_family="L1"
repeat_region
    21437..21755
        /rpt_family="Alu"
repeat_region
    21767..22143
        /rpt_family="L1"
repeat_region
    22168..22554
        /rpt_family="ERV1"

```


AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 166680)
 Waterston, R.

AUTHORS Direct Submission
TITLE Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL On Sep 30, 2000 this sequence version replaced gi:8954217.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

----- Summary Statistics

Center project name: H_NH011K18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mallo@green.nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatem, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is GS1-308H5, 200 bp overlap. Actual start of this clone is at base position 166485 of GS1-308H5; actual end is at base position 166680 of RP11-111K18.

There are polymorphic base pair differences in the overlap between the clone RP11-111K18 and GS1-308H5.

FEATURES**source**

Location/Qualifiers

1..166680

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p11.2-p2"

/clone="RP11-111K18"

/clone_11b="RPCI-11"

2..397

/rpt_family="ERV1"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/rpt_family="L1"

2138..2508

/rpt_family="ERV1"

2532..2817

/rpt_family="MaLR"

2846..2911

/rpt_family="ERV1"

2912..3043

/rpt_family="Alu"

3044..3115

/rpt_family="ERV1"

3087..3598

/rpt_family="ERV1"

3600..3780

/rpt_family="ERV1"

3781..4090

/rpt_family="Alu"

4091..4257

/rpt_family="ERV1"

4260..4865

/rpt_family="ERV1"

4902..4994

/rpt_family="ERV1"

6077..6610

/rpt_family="ERV1"

7185..7327

/rpt_family="ERV1"

7327..7539

/rpt_family="ERV1"

7579..8074

/rpt_family="ERV1"

8169..9081

/rpt_family="ERV1"

9082..9362

/rpt_family="Alu"

9363..9633

/rpt_family="ERV1"

9634..9943

/rpt_family="Alu"

9944..10150

/rpt_family="ERV1"

10151..10567

/rpt_family="MaLR"

10568..12227

/rpt_family="ERV1"

12316..12514

/rpt_family="ERV1"

12532..12582

/rpt_family="ERV1"

13068..13614

/rpt_family="ERV1"

15167..15445

/rpt_family="Alu"

16828..17212

/rpt_family="MaLR"

17548..17857

/rpt_family="Alu"

17897..18030

/rpt_family="ERV1"

18365..18652

/rpt_family="Alu"

19118..19746

/rpt_family="ERV1"

19750..19814

/rpt_family="ERV1"

19852..20169

/rpt_family="ERV1"

20224..20412

/rpt_family="ERV1"

20540..20840

/rpt_family="Alu"

20879..20997

/rpt_family="MIR"

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-128D9 is from the RP21-73 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.Chori.org/bacpac/home.htm> VECTOR: pBACE3.6.

BASE COUNT	64093 a	48521 c	50155 g	63751 t
ORIGIN				

ORIGIN

Query Match	41.48;	Score 674.2;	DB 10;	Length 226520;
Best Local Similarity	81.28;	Pred No 3	68-168.	

61 CCTCCTCCCTCCCTCCCTGTCGCCATGGCCATGAGCGAGCGCGCTCACTTGCTCGG 120

142409

Db 142408 CCGCATGCCGCTGTGGTGAATTGTCCTTAACAACCAATCCCACTCAGAGGTTTCTCCTA

181 GATCAGCGCTTGCAAGATGGCAAGCCCATGCTCCGCCAGGAGATACAGGAGACTGGAT 240

.....GCGAGCCTGGTCCATGTTTCCTGCACGAGGACTAGAT 142289

[illegible]

297 CCACCAAGCTGCGACAGCAGCTGCAGACTTCACAGCCAAAGTATGGATCGGCTCAG 356

-----AGTGTGTGATGCTGTCTCAG 142175

Db 142174 GAGATGATGGATGACCTGGCTCATCACCACATTCTCAACAAGTGTGTTG

417 ATAGCAATTACCTGNTAACTGGGGGACAGGATAAACTGCTGGCATATATGACTGAACA 476

.....GAGCAGGATTAACCTGTTCACATACACGACTTGAA-A 142057

D_b 142056 AACCTGAGCAGAACCTAACCATAATTATTCCTCAACAACCCCTTGTGGTTGCCTCGTGCTGA
|||||

98

537 GGTGCAGTGACGATAACAGATCCTTTCAGCGGATGATAAACTGTTCCGCTCTGGATC 596

CCGCGCCTGGTCTGGCATC 141937

Ddb 141936 ATACCACAGTGACAGAACTCAAATCT-~~TAA~~AAGCAATTGCGCCT TTTTTCCT |

657 ATATTCTGAGGAGAGATTGGTTACTTATGAGCATCTATTGCTTTTCATAGTG 716

.....GAGACCTGGTATGCTTTCATGTGCAGTAATTATTGTC 141819

1

1

1

1

Db	Accession	Sequence	Position
Db 99428	ATGGAACCTGCTGAC	-TGGCAAACTGCGTAGAGAAAAACATATGGCTTTGGAAATGTG	99371
Oy 1002	TGNTTCCGAGGAAGACACGCGGGCACTGCGAAAGCCAAAGATCGATTTTCAGAAACAG		1061
Db 99370	TGCTTCCGAAAGAAATATCTGGTGGCTGGCAAGGCCAAAGATGGTTTCCAAAGACAA		99311
Oy 1062	CAGAGGAAGAGCTGGCAGAAAGAAATGGCTTCAGAAATTCAGATTCATATCATCA		1121
Db 99310	GAGAAAGAAAAAAGAAATATGGCTTCAGAAATTCAGATTCATATCATCA		99257
Oy 1122	CTCTGAAAGTTAAGGCTGAGCATCAGACGTGTGCTGCGGAACAT	---ATGTTCATGG 1178	
Db 99256	CTCTCAAGTTAAGGCTGAGCATCAGACGTGTGCTGCGGAACAT		99197
Oy 1179	ACTAAACAGAGAGACAGACATGCGGCTT	CAGAGTTACTGTGCTGAGGCAAGAG 1237	
Db 99196	AACAGCAGAGAGAGAAAGCATAGAGCTTCACAGATTAATCTTCTGTTAAGGCAAAAC		99137
Oy 1238	GGCAAAAATTTGGGCAATATGAGTTAGCTCCAGTCACAGAACAGCT	-ACTCAGGTGG 1296	
Db 99136	AGCATATAATATGAGAAATATGAAATAGCTCCATTCCTGGAAACACTAATCATGCTTGA		99077
Oy 1297	CCCGTGGTGAATATGGCTAGTGTCTGAGGTGACGACAGAGAGATTGCTCACAATGT		1356
Db 99076	CCCATTAAGTGAATCTCAAGATATCAGAG	-AAGGAGGTTGAGTTATCTCTTGAGT 99018	
Oy 1357	GCCATACCTG	-CTGTTTGAATGAAAGCCCAACTTACAA-----TCTCCATTTT 1405	
Db 99017	ACATAGCCGCTGATCTTTTAATGAAATATGTAACAGATTAACATCCGACTTCTCTTAATTA		98958
Oy 1406	ACACCTAAATTTCTTTTACGTGTTATG	---TATGAGAAAGAAAAATATATGGCTTAT 1462	
Db 98957	CAATTAAGAGCTTCTTGACGCTTTATATGTCATATATGAGAGAAATATATATGGCTTAT		98898
Oy 1463	TTTTTTCGAC	-TTTCCCTTAAGAAGATGCTTTTGCTGCTTGCCTAG-----TGATCA 1515	
Db 98897	TTGTCTGACCTTTCCCTTAAGCAGATATGCTTTTGCTGCTTATGTTTTCCTTTGGTGG		98838
Oy 1516	AGAGGAGGAATATCATGATTAAGTAACCGGTTTGATCTCTTTCATTTGTAAGAGACTGCT		1575
Db 98837	TAAAGGAGGAAATCATGATTAAGTAACCTGTTTAATTTCTTCATGTGAC	---ACTGGT 98782	
Oy 1576	TCAGACAGCTG	-TATTTTGGTATCTTAATTAATATGCTCTTAATTAATAA 1679	
Db 98781	TCGGAACATCTAATTAATTTTATGTTGCTTAATAATTAATATGCTCTTAATTAATAA		98727
RESULT 11	AL645802	226520 bp DNA linear ROD 05-APR-2002	
LOCUS	AL645802/c	Mouse DNA sequence from clone RP23-128D9 on chromosome 11, complete	
DEFINITION		sequence.	
ACCESSION	AL645802		
VERSION	AL645802.11	GI:19572035	
KEYWORDS	HTG.		
SOURCE	mouse mouse.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
TITLE	1		
JOURNAL	Leonamornier, D.		
COMMENT	Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Mar 21, 2002 this sequence version replaced gi:19335832. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		


```
/organism="Carassius auratus gibelio"  
/sub_species="gibelio"  
/db_xref="taxon:101364"
```


/db_xref="GI:16768658"
 /translation="MAANNLRQIPVHLDSDICDAGYPLISACKDGP
 MLRHQDQDWDVTEGEGHGAIVNATLNTKNGADWFAWNTGAEHSPH
 KHIYKVAFDSDSENIYTGSENEKLYRVNLEOPAPREYAGHGAIKRALFEGDK
 IISAEDKTVLRMDRMNGIEYORLOFNPNPISLSIDNHLITISHSSISFEWIDTL
 KRIKEVYPTVIVASASLHPDKHVFVCGEDPKFKFDYITNENLESFSGHGPVHSAK
 FSPDGEIYASSEDEGTLRLMOTVKGKMKCTEPADLGNSSISPREVQAN"
 BASE COUNT 436 a 451 c 425 g 279 t
 ORIGIN

Query Match 20.7% Score 338; DB 3; Length 1591;
 Best Local Similarity 60.3%; Pred. No. 2.3e-79;
 Matches 573; Conservative 0; Mismatches 374; Indels 3; Gaps 1;

QY 68 CCTCCCTCCCTGTCGCCGACGAGAGCCGCTACCTGCTCGGCGACACG 127
 DB 336 CCCACTGCGGCGCGCCACMACCTTGGACAAATCCCGTACCTGACATGGCCACACG 395
 QY 128 CGGCCGCTGTGATNTGGCTTCAGGGGACATCAGCTTACGGCTACTTGTGATCAGC 187
 DB 396 CGACCCGCTGTGACCTGACCTGACGACATCTGCGACGCGGCTACTTCTCATCTCG 455
 QY 188 GCTTGCAAAAGTGCAGAACCCATGCTCCGCGAGAGATACAGAGACTGGATGGAACA 247
 DB 456 GCTTGCAAAAGTGCAGAACCCATGCTCCGCGAGAGATACAGAGACTGGATGGAACA 515
 QY 248 TTTTGGGTCATAAAGGCGTTGGGGTGCACATTTGAATTAAGATGCCACCAAGCT 307
 DB 516 TTGAGAGGACACAAAGGCGCGGTGTGAAACGCAAGCTGAACCAATGCCACCTGGCC 575
 QY 308 GCGACACACCTGCAAGCTTACAGCCAAAGTATGGATGCGGTCTCAGAGAGATATG 367
 DB 576 GCTTCCGAGAGCGCGGCTTACCGGCAAGTGTGGAATGCGGTGACCGGGCGGAGATA 635
 QY 368 ATGACCTGCTGCTAATGACATTTGCAAGCTGTGATTTCAACAGATAGCAATTAC 427
 DB 636 CACAGCTTCCAGACACAGCATGTAAGAGGCTTGGCTTCGACAGAGATTCGAGAAC 695
 QY 428 CTGNTACTGGGGAGACAGATTAAGTCTGCGCATATGACTTGAACCAACCTGAAGCA 487
 DB 696 ATAGTACTGGGACAGCAAGGAGTGTGGGCTTTCAACCTGGAGCGCCGAGGGCG 755
 QY 488 GAACCTAAGGAATCAGTGGCCACACTTGTATTAAAAAGGCTCTGTGTCAGTAC 547
 DB 756 CAGCCGAGAGATGACCGCGACATACGCGGCGCATCAAGCGGCGCTTCTGCGCGGCG 815
 QY 548 GATTAACAGATCTTTCACGCG--GATGATTAACGCTGCGCTCTGGGATCATGCCACA 604
 DB 816 GACAACTGATCATCTCGGCGCGGAGGACAGAGGCTGCGCTCTGGGAGCCGATGACG 875
 QY 605 ATGACAGAAATGAAATCTGAAATTTAATATGCTGTACAGCATGAGATATTCCT 664
 DB 876 GGCATGAGGTGACAGCGCTGAGTTCAACAGCAATCCGACAGCCTAGAGATCTCCAGC 935
 QY 665 GAAGAGAGATTTTGTATTACTTATGAGAGATCTATGCTTTTCATAGTGCAGTAAGT 724
 DB 936 GACACACCATCTACCATATCGACAGGCTCTCTGATTAGCTTCTGGAGATCGACACG 995
 QY 725 CTGGAGCCATTAATTCCTTGAAGCTCTGTGACCATCAATTCGCTCTNTTCATCCA 784
 DB 996 CTGAAGAGCTGAAGAGAGTAAAGTCCGAGCAAGCGTGGGTCGCGCAGTTTGCATCCG 1055
 QY 785 GAGAAGAGTTTCTGTGGGCGGAGAGAACTTTAACTGTACACGATGATGATTAATAC 844
 DB 1056 GATAAGCATGTGTGTTGTTGGGCGGAGAGCTTTAAGATGTACAAATTCGACTACATC 1115
 QY 845 AGTGAGAGAGTGAAGATCTTACAAAGGTCATTTGCTCCATTCACGTGTGAGATTC 904
 DB 1116 ACAGGAAACGAATGATCTTCAAGGACATTTGCGGCGGCGGACAGCGTGAAGTTC 1175
 QY 905 AGTCTGATGGGAACTCTATGCGAGGCTTCTGAAGTGGACATTAAGATTTGTGCAA 964

DB 1176 AGCCGAGATGGCGAATCTATGCCAGCGGCTCGAGAGACGSCACTCTCCGGCTGTGCGAG 1235
 QY 965 ACTGTGTAGGAAAGACCTATGCGCTGTGAAATGCGTGTGCTGAGGA 1014
 DB 1236 ACCACCGTGGCGAAGACCTTGTGTGGAAGTGCACAGAGCCCGCGGA 1285

Search completed: February 6, 2003, 14:53:09
 Job time : 6115 secs

1
2
3
4

100

100

100